9

evelopment of a Taxonomy of Human Performance:

A User-Oriented Approach

0

AD736190

Robert B. Miller

Technical Report 6
MARCH 1971

AIR
AMERICAN INSTITUTES FOR RESEARCH
WASHINGTON OFFICE
Address: 8555 Sixteenth Street, Silver Spring, Maryland 20910
Telephone: (301) 587-8201

R71-8   63

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) American Institutes for Research 135 North Bellefield Avenue Pittsburgh, Pennsylvania 15213	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

DEVELOPMENT OF A TAXONOMY OF HUMAN PERFORMANCE: A USER-ORIENTED APPROACH

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Interim Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Robert B. Miller

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1971	60	6

8a. CONTRACT OR GRANT NO. F44620-67-C-0116 (AFOSR) DAHC19-71-C-0004 (Army-BESRL) b. PROJECT NO. ARPA Order Number 1032 c. ARPA Order Number 1623 d.	9a. ORIGINATOR'S REPORT NUMBER(S) AIR-726/2035-3/71-TR6 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) BESRL Research Study 71-5

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Advanced Research Projects Agency Department of Defense, Washington, D.C.

13. ABSTRACT

A user-oriented approach is proposed for the development of new ways of describing and analyzing tasks and duties. The author considers it essential for these taxonomies to be developed and evaluated as operational information-getting and decision-making tools for use by system designers. Man-machine system design applications of this kind of tool are described in the decision areas of system characteristics, human factors engineering, selection, and training. Methodological proposals are made for the development of performance taxonomies in future years. Some current laboratory research assumptions and procedures used in developing taxonomies are criticized on the grounds that they are not adequately representative of the real world and do not lead to useful tools. Specific suggestions are presented regarding a modified laboratory approach.

UNCLASSIFIED
Security Classification

AMERICAN INSTITUTES FOR RESEARCH

WASHINGTON, D.C.

EDWIN A. FLEISHMAN, PhD, DIRECTOR

Albert S. Glickman, PhD, Deputy Director

INSTITUTE FOR COMMUNICATION RESEARCH

Arthur L. Korotkin, PhD, Director

Research on instructional, communication, and information systems and their effectiveness in meeting individual and social needs.

INSTITUTE FOR RESEARCH ON ORGANIZATIONAL BEHAVIOR

Clifford P. Hahn, MS, Director

Research on human resources, selection and training, management and organization, safety, and administration of justice.

INSTITUTE FOR RESEARCH IN PSYCHOBIOLOGY

Thomas I. Myers, PhD, Director (Acting)

Studies concerned with those behavioral and physiological mechanisms which determine performance and with those factors in the environment which affect these mechanisms and thereby affect performance.

INTERNATIONAL RESEARCH INSTITUTE

Paul Spector, PhD, Director

Research on the development of human resources in developing countries; problems of working effectively abroad; evaluation of action programs in underdeveloped countries; role of attitudes and values in social change and economic development.

TRANSNATIONAL FAMILY RESEARCH INSTITUTE

Henry P. David, PhD, Director

Research on behavioral components of family interaction; coordination of transnational population research; and strengthening of family planning programs.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Human Performance						
Task Taxonomy						
Laboratory approach, critique of						
User-oriented approach						
System design decisions						
Methodology						
Personnel psychology						
Information system						
Inventive approach						

AIR-726/2035-3/71-TR6

AD

DEVELOPMENT OF A TAXONOMY OF HUMAN PERFORMANCE: A USER-ORIENTED APPROACH

Robert B. Miller

TECHNICAL REPORT 6

Prepared under Contract for
Advanced Research Projects Agency
Department of Defense
ARPA Orders No. 1032 and 1623

Principal Investigator: Edwin A. Fleishman

Contract Nos. F44620-67-C-0116 (AFOSR)
DAHC19-71-C-0004 (ARMY-BESRL)

American Institutes for Research
Washington Office

March 1971

PREFACE

The AIR Taxonomy Project was initiated as a basic research effort in September 1967, under a contract with the Advanced Research Projects Agency, in response to long-range and pervasive problems in a variety of research and applied areas. The effort to develop ways of describing and classifying tasks which would improve predictions about factors affecting human performance in such tasks, represents one of the few attempts to find ways to bridge the gap between research on human performance and the applications of this research to the real world of personnel and human factors decisions.

The present report is one of a series which resulted from work undertaken during the first three years of project activity. In 1970, monitorship of the project was transferred from the Air Force Office of Scientific Research (AFOSR) to the U. S. Army Behavior and Systems Research Laboratory (BESRL), under a new contract. This report, completed under the new contract, is among several describing the previous developmental work. It is also being distributed separately as a BESRL Research Study.

Edwin A. Fleishman

EDWIN A. FLEISHMAN
Senior Vice President and
Director, Washington Office
American Institutes for Research

FOREWORD

The American Institutes for Research (AIR) Taxonomy Project is concerned with new ways of describing tasks and duties. The objective is to develop theoretically-based language systems (taxonomies) which, when merged with appropriate sets of decision logic and appropriate sets of quantitative data, can be used to make improved predictions about human performance. Such taxonomies should be useful when future management information and decision systems are designed for Army use.

The present report is concerned with methods used in developing these language systems. The author (Robert B. Miller), a pioneer in task analysis of performance requirements, is a consultant to the AIR Taxonomy Project staff. In this report he describes a developmental approach which is "user-oriented" in the sense that proposed approaches to task classification are subjected to several different kinds of evaluation corresponding to the interests of different kinds of applied decision makers in the Department of Defense. This user-oriented approach is presented in the context of the decisions faced by system designers, from the conception of a developmental project to its realization in an operational world. Dr. Miller makes a number of recommendations for increasing the relevance of laboratory research to the development of a practical taxonomy.

J. E. UHLANER, Director
U. S. Army Behavior and Systems
Research Laboratory

ACKNOWLEDGMENT

I wish to gratefully acknowledge the critical comments provided by my colleagues at AIR who reviewed this manuscript. The contributions of Drs. Edwin A. Fleishman (principal investigator on the project), Albert S. Glickman, and Warren H. Teichner were especially helpful.

Special acknowledgment is due Dr. Robert W. Stephenson, project director, and Mrs. Halaine Gary, technical editor at AIR, for their suggestions and revisions of the final draft.

Robert B. Miller

DEVELOPMENT OF A TAXONOMY OF HUMAN PERFORMANCE: A USER-ORIENTED APPROACH

BRIEF

A user-oriented approach is proposed for the development of new ways of describing and analyzing tasks and duties. The author considers it essential for these taxonomies to be developed and evaluated as operational information-getting and decision-making tools for use by system designers. Man-machine system design applications of this kind of tool are described in the decision areas of system characteristics, human factors engineering, selection, and training. Methodological proposals are made for the development of performance taxonomies in future years. Some current laboratory research assumptions and procedures used in developing taxonomies are criticized on the grounds that they are not adequately representative of the real world and do not lead to useful tools. Specific suggestions are presented regarding a modified laboratory approach.

DEVELOPMENT OF A TAXONOMY OF HUMAN PERFORMANCE: A USER-ORIENTED APPROACH

CONTENTS

THE USER

In developing a task taxonomy for applied problems, we must consider the user--the system designer--and his needs. In the following section we will examine specific kinds of problems which confront him and the decisions he is required to make, at least as a participant, in system development and operation. We will see specific operational decision making uses for a task classification system eventually resulting from the user-oriented approach.

The system designer, whatever his specialty, is interested in problem solving languages that help him to grasp, define and communicate the problem at hand and to create workable entities. To the applications man, a taxonomy is a tool to be tested by utilitarian criteria--criteria which generally end in some relationship between benefit obtained and cost to use.

The term system designer refers here to any individual who makes decisions about what a system including people will have to do, the contexts in which the system will have to perform, creation and selection of alternative methods for choosing and organizing the man-machine system components, and development of procedures for component interaction. The term system designer extends to specialists who provide mission objectives and descriptions, human factors specialists, and specialists in the fields of manpower selection and training, evaluation, team design and operations design.

According to my definition, individuals become system designers not merely by offering alternatives and rationales for them; they become designers when they recommend a given alternative for the system context and in doing so reject other known or possible alternatives. They make decisions within constraints on time and money that apply to system development and operation, not when they are certain they are selecting the best of all possible alternatives. Choices are made from alternatives none of which is ideal in all respects.

By definition, a "good" system designer will get better and more reliable system or subsystem performance for given development time, production and operational cost, than a "poor" system designer. Excellence in design is highly dependent on a collection of individual expertise.

Aside from bureaucratic aspects of development organizations, design is personal. For a theory to be a powerful tool, an individual mind must be capable of conceiving a bridge of relevance between the laboratory context in which the theory was developed and the specific operational milieu in which it has identifiable practical implications.

The behavioral scientist may consider the creation of a behavioral taxonomy as equivalent to theory building--the identification and naming of the sufficient and necessary variables in a general "model" of human performance. He may view taxonomy development as the creation and coding of scientific knowledge, with "explanatory power" as a major criterion of excellence, especially as applied to already published research. Decision processes in science are aimed at criteria of truth. Decision processes in system design and control on the other hand, are aimed at criteria of utility: Can the system "do the job", be built, and perform at a practicable cost?

In contrast to decision processes in science, design decisions more nearly resemble entrepreneurial business decisions. Design entails a large pattern of functional tradeoffs. What functions can be sacrificed, or to what degree, while still achieving an acceptable system product? How can we get the most (in kind and amount of relevant function) per dollars spent or years spent? In terms of tradeoffs, can the design be better (or less expensive) if a human performs the function or a machine does so? Is designer Jones clever enough to find a way to meet this step function increase in performance reliability?

A useful taxonomy for the system designer will be embedded in one or a collection of problem solving languages consisting of a vocabulary and notation for the following classes of work:

(1) Classifying events and ordering them for examination.

(2) Structuring operational problems consistent, at one or more levels, with properties of the human as a device.

(3) Perceiving workable alternatives in selection or design of components or of their mode of interaction.

(4) Perceiving tradeoff relationships in design and operations.

(5) Landing in at least a general "ballpark" of workable solutions during paper and pencil study phases of design.

(6) Enlisting the help and enlightened collaboration of various specialists.

(7) Accessing background literature, data, and existing solutions applicable to the problem at hand.

The terms and notations contained in logic and circuit diagrams are examples of "problem solving languages" for circuit designers. The notation enables conceptual flexibility with respect to the connections and values of essential electrical functions and the essential fabrication specifications in terms of component types and connections. Indeed, the circuit schematic was a powerful <u>invention</u> and from some points of view a rival in importance with the discovery of Ohm's Law.

A language, including its classificatory structures, for system or subsystem design is not an end in itself any more than the circuit schematic for a specific amplifier is an end in itself. It is a mediating tool with three anchoring positions. One anchor is embedded in the operational phenomena, the non-verbalised universe of system events, both hypothetical and actual. The second anchor is in the creative conceptual chambers of the designer's mind. The third is in the resources available to the implementer.

For presentation of an initial attempt at design of such a language based on a user-oriented "transactional" information processing approach (my current working version of a new systems task vocabulary), the reader is referred to another paper in this series (1).

From a practical standpoint, there is an intimate relationship between a useful <u>language</u> for describing and analysing human tasks and a useful <u>taxonomy</u>; they may be parts of a single descriptive procedure. The terms are used almost interchangeably in the present report.

An additional point should be made about the user of a taxonomy. He must be a skilled interpreter, or have one available. A mechanical analyst will not do. For, whatever a task taxonomy may consist of and however it may be derived, its application to the description of behavior or to documents about behavior inevitably requires human judgment somewhere along the line in linking a name to a thing or a process. The semantic act of labelling with respect to a reference is, to some degree, always an act of judgment. Conditions for the acceptable or useful semantic judgment to be exercised need to be specified explicitly.

Since task analysis and description is a professional skill, a specially trained professional person is the proper user of the task terminology and concepts to be developed. Since the use of behavioral data inevitably requires a judgment of degrees of relevance and similarity of different behaviors and behavior settings between that which generated the data and that to which the data are to be applied, professional training continues to be essential. A useful classification scheme should enable the training to be better focussed, more readily shared, and more quickly acquired than training in its absence.

We turn now to a brief look at the initial process of system conceptualization and follow with a careful examination of operational decision-making needs the system designer has for a taxonomy of human performance.

INITIAL SYSTEM CONCEPTUALIZATION

Decisions are usually made in order to solve problems and problems (especially at first) tend to be ambiguous and ill-defined. Let us consider briefly the way an ill-defined systems problem is structured and organized by applied decision makers--the process which generally precedes choices about what to build and how to operate the system.

Early system conceptualization involves some formulation of cost/performance boundaries of the mission problem and a preliminary analysis of technical feasibility for a set of mission objectives. The system problem is defined and tentatively outlined, perhaps in the form of a flowchart. Assumptions are made about the general level of personnel (skill, motivation to learn and perform) intended to operate the system, the tasks to be performed and task environments. (These early assumptions may later become decisions.) An attempt is made to identify and retrieve information about precedents for the "new" mission, job and task components--to bring to bear what is already known about the tasks in question.

In the exploration of system design possibilities and mapping of alternate roles of humans, two broad classes of questions arise (assuming a fixed population of operators): the limits of ability in some condition of load; the kind and expected frequency of human error in performing a kind of action. At each step in any procedure the system psychologist may ask: "What can the human do wrong here? What is its relative likelihood? What is its relative importance?" He would like to know how human ability and error limits could be changed by operator training or selection.

With alternate roles of human operators mapped out in a general fashion, system interfaces are sketched out and interface media between system and operator are proposed. A preliminary identification is made of contingencies and recovery requirements from malfunction, overload, etc.

Some of the cost factors and performance constraints become specific as the system takes shape. The process entails a large pattern of functional tradeoffs. After the cost factors and performance constraints become clear, preliminary estimates are made of mission success, of how much better the proposed system is likely to be than any known alternatives. If a decision is made to continue the project, the next stages in the cycle of system development require decisions of four broad types: system characteristics, human factors engineering, selection, and training. Specific decisions of these kinds are taken up at length in the following section.

SYSTEM DESIGN DECISIONS

Four major areas of design conceptualization and decision are involved in designing a system which includes people. These releted areas are human factors engineering decisions, selection decisions, training decisions, and systems characteristics deciaions (in which a system is conceived as a collection of functional requirements and entities).

These are cited in the order in which they are discussed below, rather than the order in which decisions should logically be made. Logically, decisions of the last type would be made first. It should be noted that, consistent with a systems concept, the various decision areas are definitely not independent of one another.

Human Factors Engineering Decisions

In general, human engineering is that stage of development which proceeds from a general system design hypothesis and a more or less well defined set of performance "requirements". That is to sey, a number of design constraints have baen imposed; a development schedule has been eat down, with implicit or explicit penalty clauses.

Crew Size

One practical question that may be presented to the human factors team is: "How many people are required to man the system?" This is a critical factor in vehicle systems for technical or economic purposes. For aircraft and especially spececraft the seriousness of the question is obvious--man is a supercargo rather than a payload. On buses and other commercial vehicles, the number of operators is an economic consideration that may spell a margin of profit or lose.

Before any hardware or even mockups are available, the question arises: How much can one operator be expected to learn, pay attention to, think about, and execute at about the same time? Paper and pencil flowcharts of missions may show extensive periods of time when the hardware more or less runs itself without human activity; but, inevitably there are nodes of action where many things happen within brief periods of time. Procedural redesign may seem to enable one (alert) operator to cope with nearly all of these, when viewed in the context of normal or expected operation. But troubles appear when contingencies are introduced: What if en equipment malfunction occurs? A programming error finally asserts itself? A human error is made, of commission or omission? An unusual pattern of environmental conditions occurs? This question--or series of questions--rarely leads to reliable background data on the frequency with which such contingencies can be expected singly, much less in combination.

As in many other situations of human uncertainty, when a specific rationale for action is unavailable, precedent is negotiated. Standard work crew roles are transferred from the past. Examples: pilot, navigator, flight engineer; doctor, nurse, interne; system analyst/planner, engineer, programmer; professor, junior colleague, graduate student. (No deliberate effort was made to select only trinities.) The historical existence of these groupings and roles, for better or worse, simplifies many decisions--sometimes by not exposing the decision to view. In addition, the name of the role may simplify selection and training. It implies transfer of existing personnel and training structures, for better or worse--one never knows because the alternatives rarely are investigated, and indeed it is generally impracticable to do so within system development schedules. (There is value in giving a set of job-tasks a standard role name, such as pilot. It provides administrators, as well as job candidates who already hold the title in another context, a degree of confidence that the new job can really be done and this confidence is very important.)

We have not answered the question about how the decision is made regarding the number of operators required to man a projected system. Clearly it is not, nor is it likely to be, made on purely logical and quantitative data even with the most thorough use of mission data and system information available in early development. Certainly one person cannot be in two places at one time, so this may be one determiner, although the physical configuration of the system conceivably could obviate this need. How many variables can one operator monitor at one time? How much can one operator do? Or two? "How much" refers to a large number of possible objectives that could be accomplished by the type of mission. The same difficulties arise here, although the analytic procedure may be simpler because functions that seem obviously incompatible can lead to rejection of one or the other mission objective.

Given a set of more or less abstract system operating requirements, a definitive rigorous answer to minimum crew size cannot be made. Even if the new task complex seems to be an extrapolation of those in previous systems operating in similar environments, an act of judgment coupled with faith is necessary. Formal methods for determining what differences to look for, and at least a qualitative estimate of the magnitude in performance differences, would be desirable and seem possible.

Assigning Role and Functions

Assuming a definite, but potentially tentative, decision has been made as to crew size and assuming the impracticality of complete redundancy of task skill among all members, the system psychologist is confronted with the question as to what functions to assign to each crew member. It is likely that operating requirements will impose one cut on this division of tasks. Another will be the conventional groupings of task names assigned to a job specialty or "role". Another may be what could loosely be called the maintenance of a train of thought (referred to

in the older literature as "set"). But, the questioning psychologist would indeed like to know, for a given task complex or information-handling function, how human abilities clump, and the extent to which a given degree of aptitude can overcome a given level of transferred training. (Note the three variables here--aptitude, original training, amount of training transfer.) Where large numbers of people may be involved as operators (such as pilots, computer programmers, physicians, nurses), any basis for substantive estimates could be valuable in terms of manpower and training costs. In large scale endeavors, the luxury of overselection is an expensive one, and eventually results in motivational liabilities as well.

The human factors staff may have the time and funds for modelling and/or physical simulation. Potential problem-solving benefits of a task taxonomy are described in these contexts.

Modelling

Modelling is a symbolic representation of a system, including the human components, according to hypothesized structuring of active interfaces. An interface exists where one component interacts with another component. A sample of input conditions is fed into the model--a program in a computer. The output of the model may be "time to complete the mission" if processing times are established for each component or for the transactions from mission start to mission end. Probabilities of error may be attached to each component, so that probabilities of mission success may be the criterion output from running the model. The purpose of running the model is to test a design hypothesis when it is still on paper.

On a much more informal level, a model may consist of a flow chart representing the functions and action nodes in a hypothesized system. A set of mission conditions is hypothesized. The designer traces the sequence of actions that would describe behavior in the mission, step by step, and gets a conceptual picture of where delays and extended queues will exist, and perhaps of error consequences. Obviously, a great deal of knowledge and imagination is required for this kind of modelling; but, with the right kind of talent it can be profitable in suggesting directions for modifying design.

Selecting the "right" level of description of the system's structure--the transactions among components (and level of componentry)--is critical in useful modelling. If the level of description is too gross, significant interactions will be missed and the data about the mission yielded by simulation will be misleading. On the other hand, if the level of detail is too fine for a given stage of design development, a large amount of "random noise" may obscure the major interactions. Furthermore, any consistent biases (constant errors) inadvertently introduced may accumulate into major biases in estimates of system performance.

On this latter point, a comment is in order. It is notable that when a matured engineering technology is used, and estimates of system reliability are based on known reliabilities of the ultimate components (transistors, for example), predictions of overall system reliability of the hardware are frequently underestimates of the actual hardware system reliability. This suggests--among other hypotheses--that there may be compensatory interactions among aggregates of components that, in some circumstances, may reduce individual component liabilities. The tolerance limits of one component may be somewhat compensated for by tolerance limits in another. This hunch is introduced here because the same may be true for individual human behavior. Error frequencies observed in an experimental setting may in actual performance in real life have the same tendency to occur; but, in real life there may be opportunities (through more flexible time limits, feedback checking, error tendency inhibition, team action and other mitigators) for reduction of these error frequencies or of the seriousness of their consequences. It is only, perhaps, when a system--human or man-machine or machine--is overdriven from its specifications that errors become unforgivable.

Whatever form of modelling he uses, two major kinds of information are sought by the designer: transaction time for critical segments of the mission, and probability of error. In both cases he wants to know, at the level of correctable design actions, where the excess time and the errors occur, and under what pattern of internal and/or external conditions.

These two considerations guide his selection of level of description, as well as the limits of completeness of detail in system specification and structure. At least intuitively, he will search to identify those transactions most sensitive to excessive time to complete, and those most sensitive to errors of omission or commission. Wherever bottlenecks occur because of queue buildups or because of recycling due to errors, and occur frequently or catastrophically in terms of the mission, he would like to spot them so as to provide design modifications or changes in the mission specifications.

All of these considerations relate directly to the need for a descriptive human task vocabulary and for human performance data that can be summoned and applied with this vocabulary. What level of detail in description of the man-machine-environment interfaces and transactions are necessary for even the most gross validity in estimating mission success? At present, such questions are answered by individual human factors expertise and empathy in the subject domain, or remain unanswered by default.

Physical Simulation

The system interfaces are presented to the human operator(s) in the form of displays, controls and operating environments; and task inputs to the operator(s) are simulated in order to make estimates of man-machine system performance. A ground flight trainer is a sophisticated example.

Physical simulation poses two major problems to the system psychologist. One is the degree of fidelity required of the simulated facility. This problem is related to the extent to which operator performance in the simulator is equivalent to operator performance in the real life situation, which involves factors subsumed under the transfer of training domain of human learning research. Some products of learning are highly specific to context, and seemingly trivial differences in stimulus-response relations may result in degraded transfer. In such cases, differences between simulated and real life would result in invalid predictions based on measurements in simulated environments. Other products of learning generalize more widely. Engineering technology, plus a frequently generous economic situation, has advanced so far in the last 20 years that just about anything that can be specified can be simulated physically and functionally. If cost enters the picture, the psychologist has a job to do--estimating what physical differences make for behavioral differences, and how much. It would indeed be desirable to know what tasks or task characteristics make for, or interfere with, transfer of learned skills. (This topic is considered later when we discuss training design.)

The second kind of problem posed by physical simulation is the necessary selection of situation samples, the programming of task inputs to the operator. It is impractical if not impossible to develop all the possible combinations of circumstances, external and internal, which could occur in all missions. The total circumstances, of all missions for the total population of missions to be undertaken by the system under design, cannot be known except perhaps retrospectively, and then only in theory. Some kind of stratified sampling must be undertaken for the finite number of hours available for the tests that use system simulation.

The major objective of the simulation exercise may be to uncover weaknesses in the system in order to correct them. For this objective there will be biased sampling with respect to real life representativeness of kind and frequency of mission circumstances. Input sampling will be aimed at testing hypotheses about the weaknesses and liabilities in the design. If the objective is to predict system success probability in real life, then, of course, representativeness of inputs is desirable.

In either case, the practical problem is to get the most diagnostic or predictive information with the smallest number of test samples, or the fewest hours of sampling and test. The following are the kinds of legitimate <u>short cuts</u> that an idealized knowledge of task differentiations and task structures would permit.

(a) Identification of clumps of behavior that, psychologically speaking, are independent of the rest of mission context--tasks that carry their own virtually complete behavioral context. These might be called "stand-alone tasks". Stand-alone tasks, by this definition, could be tested apart from testing the entire mission. This would enable

variations and sheer numbers of task inputs to the stand-alone tasks that would be impractical if the entire mission had to run through for each variation. Also, by definition, the results of this partial mission testing could be integrated into estimates of total mission performance on the assumption that total population variance is the sum of all the independent sources of variance. It should be noted that there are many "dead spaces" in total missions--periods when operators have little to do but wait or when behavior is neither time critical nor error-critical. (Of course, these are matters of degree rather than all-or-none; sets of total missions must also be simulated, partly to test the assumption of independence of the stand-alone tasks. Task-independence in a mission is also likely to be a matter of degree, or of probability. Thus, a continuing attempt to recover from an error earlier in a mission may encroach upon the time and attention normally available for a later task.)

(b) Identification, given a task entity, of the input variables (and the values they can take) most significant for task effectiveness by criteria of time or errors or both. This knowledge would enable study of mission conditions in order to determine which variables and values of these variables will occur, or are likely to occur. Input test conditions to the operator could then selectively (or randomly) test from these specific ranges. This would enable the test to be efficient for its intended objective.

(c) Identification of task combinations that are most likely to lead to mission vulnerability, in cases in which the operator must time-share tasks and there is variability, from one mission to the next, as to which tasks will overlap and how much. This problem is no more than a compounding of the issues cited in the previous topic. But, useful starting hypotheses could lead to useful combinations of tasks in setting up tests of capability and vulnerability.

(d) Identification of how sensitive a given task entity is to level of formal or informal training, so that appropriate requirements can be imposed on the human operators selected as subjects for the simulation exercises. Bare mastery of a task may, in performance context, yield quite different results both qualitatively and quantitatively than the same task that--in conventional terms of learning psychology--is "overlearned". If the mission is performed by a cooperative team of operators, the level of team training is obviously significant as a basis for selecting "representative" subjects.

(e) Identification of the distribution of individual differences among highly practiced operators, who are likely to be representative of the population of operators of the system in full-blown operation, on the task entity in question. We are citing ideal knowledge, and not limiting ourselves to what at the moment seems to be a practical (or even theoretical) expectation. If this kind of knowledge were available, it would help determine the number of operators to sample in test

exercises. This matter has, of course, circularity. Criteria used in training and operations may have well-defined cutoff levels on what is acceptable performance.

Of the five topics cited above, probably only the first two or three should be expected to have useful solutions in terms of data in a behavioral data bank. The massive costs of system simulation justify considerable effort toward finding "modules" of behavior that can be tested in isolation from the context of total mission exercises, but which are predictive of mission success. A module, by definition, is a unit of structure whose behavior, in any range of conditions under which it is expected to operate, is predictable in terms of function and, therefore, whose behavior parameters can be exhaustively measured.

Selection Decisions

Selection vs. Training Alternatives

"Shall we select from operators who have similar skills and abilities and retread them with new skills, or will we do better to select new people and train them from scratch?" In the past, this question may well have been asked about maintenance personnel with troubleshooting and other maintenance experience in mechanical gear who were shifted into maintenance of electro-mechanical gear in tube technology, and again with the shift from tube to solid state technology. In most cases, the decision is more heavily weighted by practical factors, such as what to do with an obsolescent manpower pool, than the respective values of selection vs training per se.

Two major questions appear in selection vs. training. One is the degree of expectation that the potential capabilities (aptitudes) of the incumbents in the old task are adequate for learning and performing the new tasks. A higher, as well as different, cognitive capability may be required for electronic troubleshooting than for mechanical troubleshooting; if this is significantly true, then a new subpopulation of recruits should be searched. On this consideration, the personnel man would like to have some objective measures--or library reference information--for determining how much more difficult Task A^1 is than Task A^2, in terms of aptitude. He would like to be able to make a decision but save the time and cost of empirical trial and error.

The second major question in selection and training is: "What is the training cost (measured in time to learn) of Tasks 1 ...n that comprise the new operator position?" This estimate can be balanced against an estimate of cross-training costs.

The reference image of task which the reader has in mind should not be restricted to routinized repetitive activities, but should extend to more complex cognitive activities such as the interpretation of an unusual

pattern of cues or of a garbled message, managing a tool that fails to perform as it did on previous occasions, deciding whether or not an emergency exists, holding in mind half a dozen messages at the same time, searching out a route through crowded channels, remaining alert for near-threshold cues in a large ill-defined scan area. Every high level skill must on occasion be exercised in at least several of these circumstances.

A brief examination of a situation in which the issues of upgrading and cross-training versus selection and training from scratch have significance, is in order.

Traditional operators of computers in the past were often regarded as low level technicians. Only rather casual selection criteria have been applied; and, computer operators seldom receive more than a few weeks of training before they are put into the job. In traditional batch operations of computers, a backlog of jobs is stacked and a job is done when it is finished; stringent criteria are rarely applied. The consequence is a high variability among operators (and installations) on system throughput. These conditions have been accepted as tolerable, usually because the recipients of computer output accepted the work if delivered within rather loosely defined norms.

But the computer customer is on the threshold of a radical change in the interaction of computer user with computer installations. Shortly there will be large numbers of terminals connected to central facilities, and users at these terminals will want their answers at once. The computer operator will have to be able to respond within seconds to any of a large family of contingencies. He must assist in managing a large and complex traffic, and intervene when a situation has not been anticipated by an automatic program--which probably will occur frequently. He will have innumerable control choices, and he will have to anticipate their consequences. There are many thousands of traditional computer operators. It would be an organizational convenience to retrain them for the new kind of operator job. How many and which ones are salvageable, if any?

We have again the same set of problems that appeared in the decision on whether or not to train mechanical maintenance personnel to be electronics troubleshooters: amount of transferable skill from the old to the new job; differences in aptitude, if any, required for the new job; differences in work interest. The latter is perhaps a more crucial factor in the decision about the computer console operator. It seems essential that he be motivated to give service--be interested in serving people. The need for this interest, in addition to job skills, can be demonstrated in a variety of ways not, perhaps, immediately relevant to this line of discussion. If, however, the work habits and job attitudes of Job A are contradictory to those of Job B, the interference effects from shifting a man from Job A to Job B may be more pervasive and far more persistent than the interference that arises because of differences in skill requirements.

Traditionally, personnel decisions involving large numbers of people have been swayed more by "who is available" than by rational selection opportunities and procedures. But the costliness of simplistic expedience coupled with what is more important when the system is running (human exasperation with bad service) is likely to compel greater effort among ethical decision makers. That effort might be more fruitful if at least a rough checklist and gross tradeoff picture of the variables could be provided to the manpower specialist, personnel specialist or consultant.

As is well known by those who practice in the field, years of empirical study are never available for exploratory development of good design hypotheses, although months of testing and refining a reasonably good starting hypothesis can sometimes be arranged. A high batting average on good hypotheses demands some combination of practical expertise and working theory that can make the most from incomplete data about task requirements, variability among applications, tentative system specifications and environments, and uncertainty in the administration of personnel policies.

Here, again, we see the need for a conceptual schematic that can help to structure what exists as a class of unstructured problems--or in fact as a concrete case of an unstructured problem. When the best that can be expected are informed guesses, a family of coordinate variables would help to make better guesses. Procedures for using the conceptual schematic should not lead to the kind of bookkeeping, counting and detailing that can quickly obscure the dominant central issues.

Skills transfer is a factor that would be useful for the systems psychologist to know about in choosing between selection and training tradeoff levels. Research has revealed a great deal about specific negative transfer of training effects (e.g., transferred letter positions in nonsense syllables between the first and second lists learned), effects which in the practical world are generally transitory. (There are a few exceptional, dramatic cases of toggle switch reversal that lost an airplane.) But, little indeed seems known of what general effects are learned and transferable from a long practiced task-context to new learning problems in a similar task-family. ("Similar", of course, must be defined here as what is transferable.) What kinds of cognitive schematics or maps of environments, cause-effect relations, identifications, procedural strategies, expectancies and other kinds of mnemonic supports for new learning are acquired and transferred? Can they be identified even grossly without a massive empirical study which, in any event, must always be oriented to the single case?

Notice that we are not now examining any single factor in isolation. It is relevant here, but too limited in scope to ask the question, "Is there any carryover of an acquired skill in diagnosis from mechanical to electronic equipment and from these to medical diagnosis--or in the

reverse order?" The best examples could be provided if the answer to the question posed were already roughly known, and could thus be sampled in specific cases. Some broader questions might be asked.

What can be expected to transfer (in the way of savings in training) from the skilled operation of any vehicle (such as an automobile) to any other vehicle (airplane, earth moving tractor, submarine) aside from the specifics associated with such controls as accelerator, brake and so on?

Why (aside from motivational factors) would we expect that a skilled professional pianist would transfer his skill more quickly to learning the violin than learning to type? Or would we? In this case, the ability to read music from printed notes, hold the information in his head (the concept of the sound represented) briefly until it can be executed by the fingers, is clearly a transfer mediator. The typist buffers sets of symbols clustered in the form of words and phrases--a seemingly quite different kind of pattern and content. But the finger motions of typing are more similar to piano playing than to violin playing!

Obviously, the difficulty of imposing controls in studies of this kind are enormous, depending on what one needs to find out. Introspective observations by the subjects are apt to be worse than useless for drawing conclusions, although potentially useful for developing hypotheses about what to observe and to validate.

Career Path

Organizationally, a career path may have little more meaning than a route through a number of status positions that move from the bottom of an organizational chart toward its top. More significant, psychologically, is some basis of career founded on transfer of training, attitudes and interests that progress from lesser to greater competence on the one hand, and lesser to greater value to the enterprise on the other. Traditionally, the competent worker moves from peer member of a work group into management. There may be psychological and organizational justification for such policies among unskilled, semi-skilled, and skilled labor in the old manufacturing context. But among a large variety of professional people, the transfer from technical (in the broadest sense) skill to managerial responsibilities threatens the individual and his organization. More lip service than practical recognition is being given to the problem by some companies that have large numbers of professional people, and depend on technical talent as well as managerial talent.

There are other pressures involved in personnel policy decisions that have bearing on a better understanding of career paths--what they should be and can be. This is the humanistic philosophy of Maslow, McGregor, Argyris, and others who advocate "self-realization" in work. At present these concepts seem semi-mystical, but they do reinforce interest in the

question, "What is a psychologically meaningful statement of a career path, and how, in the concrete case, does one chart such a path early enough in a career to make any practical difference?"

Undoubtedly, native aptitudes and transfer of training factors interact, but at what levels and in what ways are obscure, unless the personality studies on leadership (i.e., McClelland) and those on creativity (i.e., MacKinnon) have general applicability.

In broad terms, we would like to know what a given pattern of experiences. associated with the set of tasks comprising a position or job, will enhance in the form of aptitudes for "new" tasks. Analytic handles, even for starting one investigation with hypotheses suggesting what to observe and measure, seem lacking in this enterprise. In terms of the general theme of this section, we may ask, "From a transfer of training standpoint, what is a 'task'? From an ability standpoint, what is a 'task'?" And then we may ask, "From an individual's and organization's viewpoint, what is a psychologically justified 'skill enlargement' in the concrete case?"

Selection Criteria and Objectives

Ideally, the personnel psychologist would like to be able to examine a proposed or actual set of job-task behaviors and/or requirements, and parse them into a collection of subsets each of which would point unambiguously to one of a comprehensive, standard collection of selection tests, and thus compose his selection battery. The only empirical need would be to jockey around a few weight values in order to finely tune the overall criterion.

In order that this repertoire of tests would be reasonably small in number, yet comprehensive and have useful validities, the tests would have to be relatively free of task content and represent instead the task structure or operations. A minimum dependence on task content is probably the objective of virtually all test developers concerned with general abilities and aptitudes. Unfortunately, since the human is an associative mechanism (in the sense of "stimulus and cognitive associations"), the distinction between task content and task structure is (unlike inanimate mechanisms) not generally self-evident. Researchers (e.g., Guilford, Fleishman) spend substantial parts of their professional lifetime in painstaking development of "pure" tests--tests free of content or context, and independent of what is measured by other tests.

Whether selection tests are to be developed or whether a choice is to be made from a repertoire of tests, it clearly would be desirable for the system psychologist to be able to specify the structural nature of the key tasks in a new job and to provide a useful range of content samples (assuming that a set of analytic and descriptive concepts were available for distinguishing the structure of the process from the content).

Since the structure of a task must, to a large extent, be bound up in the procedure by which it is performed, assumptions about task structure should stipulate training operations that are consistent with the operator learning the assumed structure. This would control the random variance, as well as genuine bias, between task specification, selection and training procedures.

If the structuring of training--the segmenting and sequencing of training content--could be made compatible with the task vocabulary and the selection test vocabulary, a systematic way of improving all three by continuous adjustment would be evident. The hypothesis would run as follows: Whatever can be learned as an independent chunk and then inserted into the other chunks of what has been learned, with little or no interference effects when the chunks are put together, represent independent abilities. Although this statement clearly needs some refinements and qualifications, it has tempting promise as a basis for theoretical and practical objectives.

Considering the distinction between task structure and task content, an information processing approach to task analysis, such as that suggested in general terms by this author (1), would seem to have the most promise.* This approach enables flexibility between the structural relationships of the operations and distinctions between these structures and task content. It would seem also to have most value to the system psychologist who has to project task behavior from system blueprints and verbal statements. This flexibility could, of course, also be its major liability in complicating the choice of appropriate level of description both for training and for selection test decisions.

Training Decisions

There are three major criteria for training programs: relevance of what is learned to job tasks; efficiency in the total learning operation that leads to on-the-job criterion performance; completeness in learning, to the degree necessary, all of the tasks that comprise the mission--including responding to contingencies.

Task information and task reference information directly support all of these criteria for good training, in theory if not in general practice. Unfortunately, the formats and procedures that have been widely practiced as "task description" often seem to miss the heart of the task examined, either as performance requirement or as behavior. Exercises in clerical diligence have often substituted for insightful analysis.

---

*This distinction is not necessarily recognized by many researchers who seek to characterize information processing parameters and models of behavior.

The competent training designer wants task information in order to direct decisions about performance criteria, training costs, part-tasks, errors, procedure design, and training simulators.

Performance Criteria

Performance criteria are specifications as to what response is demanded of the operator to what range of input conditions and environmental states. If complete, performance criteria include the range of contingencies and failure conditions from which the operator is expected to recover. These include task complications created by the operator's own errors.

Training Cost Estimates

Estimates of training costs, especially in terms of time to learn to acceptable performance levels, may have decisive effects on some projects, and be another basis for the choice of system alternatives. If tasks that comprise a new position, or segment of it, can be characterized in such a way that reference material can be used for even coarse predictions (25% error or even more might be good enough), these decisions could be made earlier in development.

The expression noted above, "learn to acceptable performance levels", is significant. Formal training does not generally take the operator to acceptable performance levels on the job; usually this is completed by on-the-job "experience". Because of this practice and the understanding of training as formal training, estimates of "training" time may be quite arbitrary. On-the-job training is subject to great variability in effectiveness and efficiency, so that in many cases where time-to-learn data might have been kept, it would be relatively useless for predicting total training time in a total training program.

The training philosophy may posit that the program is responsible only for "teaching the student how to learn the task" rather than be responsible for task-directed training with added responsibility for checking out students with a task performance criterion level acceptable for mission performance.* In the first case, task information is irrelevant to training content, as is the training content likely to be to the task.

---

*Obviously, the training of astronauts is <u>not</u> an example of this philosophy. Much industrial training, and in general the training of service personnel (e.g., maintenance), does exemplify it.

There are growing needs for estimating training time, even without pilot studies. Retreading and cross-training in industry and in military and other government positions may affect large numbers of people and many manhours of investment.

Part-Task Training Segmentation

There are many reasons for splitting the activities of a total mission into independent, or relatively independent, training segments. One is physical cost. For example, it is less expensive to train on a part-task trainer than a full simulator. Another reason is efficiency. On a part-task trainer the student can be exposed to hundreds of input variations in the time that a single mission may be run and which may contain a large proportion of what, for training purposes, is dead time. The part-task trainer can be responsive to individual differences in rate of learning one task versus another.

The well-known liability in improper part-task training is that it may be largely a waste of time. If the task modules are incorrectly chosen, or used at the wrong stage of learning, there may even be persistent negative transfer when the part-task is performed in full mission context.

The expression "part-task" is used here for convenience. More properly it should be called "task" training, in contrast to "mission" training. A significant dimension of utility for a task description methodology would be that of pointing to well-defined segmentations in training schedules for skill development which could be validated experimentally. If the task was sometimes time-shared with other tasks, but not always concurrent with them, then almost certainly another set of considerations--stage of learning--would intersect that of task differentiations. For instance, as a time-shared task approaches the stage of behavioral automaticity, it would seem desirable to start practice in full work context in order to force development of the appropriate time-sharing mechanisms in attentiveness.

A substantive example of training segmentation can be cited. In this case we intend to train a troubleshooting ability on a piece of gear. Should the student learn the various procedures associated with making checks at various test points at the same time that he is learning the cognitive skill of deciding what next test point to check, and of the inference-making based on the last check plus those that preceded it? Common practice lumps both activities together. The consequence is relatively few total exercises in troubleshooting, and an underdeveloped cognitive skill. Assuming some degree of independence between the cognitive skill (deciding and inferencing) and the procedural skill (making tests), it would seem desirable, at least in early stages of learning, to separate practice on the test procedures from practice on the deductive skills. The latter would require only symbolic representations

of the system being diagnosed and of test values revealed by the student's selection of a symbolic checkpoint.

In these terms, exercises could be graded for difficulty, if the training designer knew explicitly what the student should learn--in this case one or more cognitive strategies. The training designer must know what the structure (or strategy) of the cognitive activity should be if it is to be taught at the symbolic level. Trial-and-error behavior in a symbolic representation will produce no more skill than trial-and-error behavior in the real work context, and the latter may have the advantage of enabling the operator to take trial-and-error short cuts.

Both task structure and task content obviously are essential for training design purposes. The paragraphs above demonstrate the point, as will the sections that follow.

Kinds of Errors to be Expected

Task description, and the reference information it might summon, should enable the training designer to anticipate error tendencies in learning and in performance. In many cases, knowing about the nature of these error tendencies in advance enables at least two corrective actions to be taken. One alternative is to program the training so as to expose the tendencies and provide both feedback and practice opportunity to correct them. Another alternative is to design the task procedure, situation permitting, so as to counteract the error tendency. Still another possiblity is to create one or more kinds of habit redundancy the combination of which will increase dependability of correct response. (This latter procedure may also be useful in training for rarely used task capabilities that are subject to forgetting.)

In many cases, the kind of error tendency is as much a function of the behavioral context in which the task is performed as of any elements in the task itself. This is especially true in situations in which short term memory must hold variable information for later response, through periods of distraction by other activities.

An example of a characteristic error tendency of significance to training would be to the point. Observations of troubleshooting behavior in a wide range of situations frequently show a strong tendency for the troubleshooter to generalize a diagnosis from one occasion to another. The same effect may arise from a variety of causes. But, before the troubleshooter has made enough checks for a logical justification of the cause in the new failure condition, he leaps to the conclusion that the cause is the same as what he found to be the cause in previous trouble in which at least some of the external symptoms were the same. The more dramatic the past occasion, or the more recent the precedent (or the more frequent in the troubleshooter's experience), the more powerfully it leads

the troubleshooter to disregard conflicting or contradictory evidence and to persist in what may be an erroneous hypothesis. Recognizing and anticipating the phenomenon, it becomes relatively obvious how to program a series of training exercises to eliminate or at least counteract this tendency and, through practice, teach the troubleshoot to act logically. (The phenomena reported in the general literature on "hypothesis formation" are, incidentally, relevant to the troubleshooting task, both for anticipating error types and for successful structuring of the task itself.)

Designing Work Procedures

The training analyst and designer may receive his training specifications with task procedures already spelled out by the system designers or human factors specialists. If not, he may be challenged to tackle the job himself. This requires that he have as much task information as he can get. If he were guided by a systematic conceptual scheme of general task "structure", he would know what questions to ask about behavioral context in the mission which he could relate to behavioral alternatives in designing a procedure to be learned and remembered with relative ease and reliability. This learning would compensate against expected error tendencies.

Kinds and Degrees of Simulation for Training

The points made in our earlier discussion about physical simulation apply here. Although fidelity in copying the physical work configuration of the operator may often be trivial in cost, representing the dynamics of situations including feedback dynamics can be highly expensive in time, dollars, and effort. Lacking a training establishment with virtually unlimited time and funds, task information that points to references which can guide, even within gross practical limits, the sufficient requirements for effective learning and transfer of learning, will make a substantive contribution.

Procedural tasks that, at least early in learning, use a large amount of verbal mediation can predictably profit from practice on relatively crude, semi-functional mockups. The major liability is a lack of enthusiasm on the part of the instructor which, of course, readily transfers to the student. The motivation-incentive picture can change when the student becomes preoccupied with learning the task operations and ceases to be preoccupied with the object on which he is practicing. As anyone watching children or adults engaged in games will quickly realize, realism is a state of mind. This psychological knowledge is highly relevant to training operations.

Task and mission information will direct the program content of the trainer and may be modified according to whether the intent is to train or to test for predictive purposes.

Accessing the Training Literature

The names given to task or to behavior ensembles that make up task performance would, ideally, be names not only descriptive of performance requirements but also that would reference the relevant literature. The objective in using the literature is to increase the efficiency of training in reaching some given level of dependability and competence. In some cases, training may raise the effective ceiling of the human performer; apparently this is what some of the great coaches in athletics and the arts are able to do.

This assumes that there is, in fact, literature that is relevant to the task and behavioral configurations of interest, and that it is reasonably accessible once identified. Making this assumption, it should be possible to obtain good hypotheses for training technique if the task descriptors could be matched with the research descriptors.

It should be noted that although valid on its own grounds, the research literature may be misleading in a real work environment. For an excellent review of this issue see an article by Chapanis (2). See also the critique of information processing models by Reitman (3), and on research methodology by Bakan (4).

Three major areas of training interest in which research findings could be of most significant help can be identified. One is the kind of conceptual training that is most effective in learning and performing the task. A second is the kinds and orders of information feedback to deliver to the student, perhaps varying at different stages of mastery. A third is guidance on situation sampling and progression that combines effective learning and transfer of training from the school to the work situation.

Systems Characteristics Decisions

Manpower Estimates

Large establishments in the governmental and industrial spheres have bodies of manpower with job codes and skill descriptions from which selection must be made for new or seemingly new positions and position-tasks. Although extensive research has no doubt been conducted in the attempt to find a skill nomenclature that can be matched to task description nomenclature with transfer of training validity, it is likely that conclusions remain tentative. It is possible, perhaps likely, that only gross matches between task requirements descriptors and human skill code descriptors can ever be made. But, after determining what the practical limits might be, it would be worthwhile to try to achieve them. The cost of skilled manpower is not likely to be reduced in the future.

A skill is generally a class name with a task reference and a content or context reference; these references may be explicit or

implicit in the name and definition of the skill. (Historically, skill nomenclatures lack semantic discipline. But, this may merely be indicative of the fact that task description lacks semantic discipline.) Conceptually, a skill may rest primarily on an aptitude base, or it may rest primarily on a training and job experience base. Practically, of course, a skill rests on both. A thoroughgoing manpower management procedure should probably identify its assumptions in this regard.

In brief, the personnel psychologist would like to receive from the system psychologist task information about a new enterprise that would enable translation into an index for selecting manpower skills effecting the best compromise between availability and amount of training time for the new position.

Performance Monitoring

The management of a system in operation has the need to control its behavior. This implies measuring its performance against reference standards, detecting deviations exceeding tolerance limits, diagnosing the correctable cause, and taking ameliorative action. This generality applies in particular to the human operators in the system, and the "evaluation" of their behavior. Evaluation is meaningless without some kind of reference and reference operation.

The task definitions and the task requirements provide at least one major dimension of reference in monitoring the behavior of humans in systems. Furthermore, when deviations occur, management has an explicit reference for analysis of the trouble down to the minimum correctable behavior to be modified--a criterion of efficient control.

Task description provides management with an objective language for communication with the operator; description can be substituted for value expressions and resentments they characteristically arouse.

In short, with suitable descriptions of operator tasks, system management is in the best position to effectively and efficiently monitor and interact with its operating personnel in achieving and maintaining the performance for which the system was designed and checked out. In addition, the most objective basis becomes available for perceiving where the original design specification is inadequate or obsolete, and for pinpointing where changes are essential in procedures, components, incentives, or objectives.

Insofar as rational behavior is expected and desirable in systems, including those with human components, a task-reference is essential for control and the communications required for control.

Selecting Competitive Revisions of a System

A system complex has a generation of life; it is installed, matures, and then, inevitably, competing new generation systems are proposed or enter the lists as competitors of the older system. An inevitable question arises: "What does the new system do or do better than the old one?"

If available, task specifications and actual performance data associated with task specifications can be significant or crucial in making key comparisons among competing systems, or between the old and the new. Samples of actual "mission data" could provide information about environments and environmental effects on various task performance--errors, overloads, short term and long term learning effects.

Rarely is such information available or interpretable in a way that would permit such hard-boiled comparison to be made. As a consequence, a system's management substantially lacks the foundation for specifying what it can confidently expect to be an improved version of an existing system (assuming it meets the new specifications), nor can it make cost-performance evaluations in regions of overlap between competing systems.

A consistent and more or less standardized technology for describing tasks and task requirements would have to be applied to all members of the competing group of systems in order that comparison data could be evaluated from a common base. Hopefully, such a task description technology would enable interpretation of differences in task variables.

And, if the task description technology were applicable not only to human behavior and performance but also to system behavior generally, including the inanimate portions insofar as they were information processing operations, a very great boon indeed would be given to comparative system evaluation and choice.

We have now come full-cycle in the evolution of a system life-cycle from starting concept to senescence and regeneration. Descriptive records of failures and successes associated with past experience do not in themselves guarantee that past failures will not be repeated in new cycles; but, without such records, repetition in future enterprises can only be avoided with good fortune.

We turn now to a discussion of the need for an empirical-inventive approach to development of a performance taxonomy (or language) and a review of several promising approaches of this type, having considered in some detail the usefulness of just such a tool in structuring system design problems, defining variables, highlighting decisions and action alternatives, and recognizing trouble spots. An important step in the work of developing such a task taxonomy is that of defining a set of conceptual objectives which specify its intended applications (e.g., applications to the system design decisions we presented in preceeding pages). As we shall see, these objectives suggest useful criteria of a quantitative nature for evaluating a taxonomic product.

BLANK PAGE

AN EMPIRICAL-INVENTIVE APPROACH

In analyzing the development and evaluation of a methodology for task description and analysis, we borrow a principle found useful in other problem solving contexts: Establish operational objectives; let the objectives guide motion towards the solution and constitute the criteria for measuring the success or acceptability of the solution.

Conceptual Objectives for Use in Evaluation

One set of conceptual objectives for a task description and classificatory methodology can be characterized by the applications intended for that methodology. Examples of such applications, grouped in terms of system design decisions, were described in the previous section. Those described at length in the text and those relating to early system conceptualization and the organization and structuring of an ill-defined systems problem are summarized below. Individually and collectively, these areas would form the problem set for testing the utility of a performance taxonomy in the personnel subsystem. The diversity of these problem areas suggests that a taxonomy may have to consist of more than one dimension or subset of terminology. This would, of course, be less desirable than a single set.

Early system conceptualization. Preliminary analysis of technical feasibility; formulating assumptions about the mission problem and general implications of personal environments, task environments, class of personnel, etc.; search for general reference information.

Structuring the unstructured problems at the system level. Formulating system tasks, mission definition, general alternatives for role of human(s), control nodes in the mission process sequence, general sketch of system interfaces; preliminary identification of contingencies and recovery requirements from malfunction, overload, and so on; archetype operation and system design possibilities.

Human factors engineering design and test. Deciding size of crew required to operate and support the system; paper-and-pencil modelling of behavior in a mission in a tentative design; anticipating error liabilities of operator, pinpointing critical operations, estimating human load qualitatively; reference literature search for guiding design principles and performance data relevant to task; physical simulation--requiring fidelity, situation sampling including time-shared streams of activity, interpretive projections of simulation data.

Selection versus training alternatives. Estimates via rational analysis of equal or greater level of aptitude in new task-job to reference task-job; estimates of training costs versus retraining costs (magnitude of transfer of training from reference task), including job experience as an estimated training cost factor.

Preceding page blank

Career path. Psychological picture of growth of task skills as consistent extension of learning; picture of growth and expansion of skills as the fulfillment of aptitude potentials; contrasting but supplementary points of view about extensions of human capability.

Selection criteria and choices. Task analysis as a statement both of a processing structure and a processing content, with selection testing aimed at determining essential human processing structures with content only sampled; glossary of task names and definitions matched with glossary of selection tests.

Training. Performance criteria; training cost estimates; part-task training segmentation and sequencing; assessment of kinds of error to expect and development of practice programs to mitigate these errors; design of work procedures for efficient learning and effective performance; determination of kinds and degrees of simulation and use of standardized task modules; accessing the relevant training literature; estimating transfer of skills and abilities from existing manpower pools.

Performance monitoring. Task requirements as basis for monitoring human behavior in the system environment and as objective subs''tute for personnel evaluation and its liabilities.

Selecting competitive revisions of a system. Generation of systematic records of system-task experience for specifying requirements of a revised system or for examining and comparing proposed new systems with performance and cost/performance capabilities of the old system; use of archival records as reference data.

The Need for an Inventive Approach

When a problem is well-defined, the invention of a tool can be quickly accomplished. The need for a practical problem solving language, including a taxonomy, is immediate and increasing. The proliferation in variety of man-machine "information systems" is evidence. The cost and sophistication of these systems demand more than hit-and-miss developmental technologies. But the timetables of scientific research are not geared to the pressures of developmental schedules. Even if scientific development of behavioral taxonomies yields products of practical utility, the present state of affairs indicates that the date will be many thousands of pages of controversy and many years distant.

Inventions can be adopted, improved, discarded as needs and knowledge change. Ideally, the invented taxonomy would have parallels in research so that discoveries in the laboratory would supplement or modify the instruments used in applications; qualitative variables might become quantitative parameters.

An empirical-inventive approach is one which combines invention, test of the invention in applications, and modifications in the light of these tests. By definition, it is impossible to characterize any complete universe of inventive approaches.

Several examples of such approaches, that hold promise in one or more ways, are described below. Certainly they are not mutually exclusive. In principle, they all combine background data and individual or collective expertise applied either to (a) better techniques for making and using task analysis for reference and design decision purposes or to (b) making actual design and/or performance hypotheses which are subsequently tested and validated or corrected, including the data base of reference information.

Personnel Subsystem Decision Matrix

The structure of a given class of decisions can specify the kinds of information necessary and sufficient for the making of a decision in that class. One basis for structuring a class of decisions is the repertory of alternative choices available to the class of problems in which decision choices must be made. It is unnecessary to postulate mechanical means of arriving at decisions for these principles to hold; the decision may be reached subjectively with judgment and under uncertainty.

A set of selection alternatives, for example, might well include the decision to select on the basis of an existing ability to perform the task or some psychological equivalent of it based on transfer of training estimates. The candidate population might consist of assessments of aptitude or ability to learn the reference task or job. Or the choice of selection route might consist of the alternatives of: teaching a problem solving skill by means of extensive drill in rote procedures; teaching the same skill by way of concept and principle, where the operator deduces the specific response to a specific situation from a general principle. Another testing strategy might consist of selecting by test only for certain necessary capabilities on the basis of cutoff levels for rejection, and test the remaining candidates by directed training exercises on factors relevant for rejection.

These are thoughts about an approach to structuring the selection procedure with decision alternatives. Notice, however, that the examples (at least as chosen) interact with training decisions. It is true that one class of decisions can be functionally related to another class of decisions by the fact that one set has tradeoff factors for the other set of choices. A simple example of such a tradeoff: a lower cutoff level on the selection procedure <u>may</u> be compensated by more extensive training. A larger standard error of estimate in prediction from a test may be compensated by more extensive training plus deliberate attrition during training.

These examples are not intended to imply that a personnel subsystem decision structure exists today. It still remains to be invented, and its invention probably will have chance of success (utility) to the extent that it is tackled systematically and not randomly.

Aspirations for precision and a tightly interlocked qualitative/quantitative pattern of relationships should be continuously viewed in the light of operational common sense. It is only when the operator is working under maximum psychological load in an environment that is unforgiving to mistakes made either in timing or in kind that a finely honed predictive apparatus makes any sense at all. In most tasks most of the time, human limits in performance are not even in sight, so that motivational-incentive factors are far more significant to level of performance than skills and tools and work space design or niceties in selection and training background. This point in fundamental practicality is apt to be overlooked by classical researchers. It is, of course, well known to civil engineers, electronic engineers and package designers who, after estimating some minimum construction requirement to some estimated maximum stress (or probable maximum), use a safey factor of from two to fifty in construction. The degree of precision that can be useful in the estimate is directly related to the size of the safety factors to be permitted in design--inevitably overdesign, at least in some regards.

There are statistical treatments that can make the addition of "safety factors" relatively precise with respect to risk, if the parameters of the environment and of system performance are precisely known. But the constellation of factors that may at some time all work together to tear a bridge apart cannot be known, and even if they were, building to all such "worst possible cases" would be impractical--nothing would ever get built because cost would be prohibitive. Safety factors are based on predicted <u>estimates</u> of system load and <u>estimates</u> of integrity of building materials and construction procedures.

Furthermore, the notion of a system "steady state" can be a misleading metaphor taken from mechanics, usually supported by a short range statistical view. The complex system which contains human components is continuously changing, exploring properties of the environment and properties within itself as manifest by transactions. It is continuously "learning", although at different rates at different times and places in the system. Adjustmental changes tend to lead to structural changes. Thus, a system tends to be a conglomerate of learning entities. The significance of this by no means original observation is that a decision-making matrix for the personnel subsystem does not have to aspire to "perfect" predictions and "perfect" personnel subsystem design decisions. Like the engineer's first design hypotheses on paper and later engineering model on the bench, they should be reasonably good. But a substantial part of design technology consists in knowing how to make what is an inadequate first pass at design into a better one--how to steepen the slope of the learning curve for that system enterprise. In large part, this consists of setting up test models

so that measurements appropriate to options for improvements are easy to obtain, under conditions relevant to actual system performance.

This observation has a central significance to the invention of a useful decision structure network for a personnel subsystem. The structure should enable flexibility in modifying a set of decisions (e.g., decisions in selection, procedure design, work space design, training) as the system is developed. Data format structures should help to capture information about the properties the system itself is developing as well as about the environmental conditions experienced. In brief, the design of the decision structure should foster a steep learning curve for the personnel subsystem within the development maturation cycle of a given system. It is certainly easier to invent a heuristic structure to such principles and objectives than to attempt creation of a perfect Cassandra. Technology should be concerned with the art of the possible.

In brief, identification of a class of decisions will specify the kind of information necessary to make a reasonable choice. In other words, the structure and nomenclatures in a task analysis technique would be derived from a personnel subsystem decision structure. In theory, both the format and kinds of content necessary and sufficient for describing task operations and environments in order to make these decisions could be logically derived from an explicit decision structure.

Practices of Experts

Task analysis has been used for nearly twenty years for a variety of purposes and in a variety of forms. It should be useful to interview in depth, on a more or less clinical basis, several score of the more active professional "getters" and users of task information. Such an effort would provide information on how the practitioner performed a "task analysis", his dissatisfactions and successes in referencing the "literature" (including his own knowledge background) and, perhaps most important of all, how the task information he derived was used in personnel subsystem actions and decisions. It would be useful to obtain detailed estimates of the kind and amount of information the investigator set down in contrast to the amount and kind he carried in his head. An attempt should be made to determine the key concepts used by each practitioner. The intent of the study would be to profit from diversity rather than deplore lack of standardization.

A questionnaire should not be used to collect these data. The work requires a competent and patient interviewer with broad knowledge of human factors, training, and selection, and someone with good operational sense as well. Symposia in which participants contribute technical papers on "how and why I do a task analysis" would risk defeating the purpose of obtaining shirtsleeve descriptions of what really happens and what is really done, because of inevitable tendencies towards impressing colleagues and "originality".

Preliminary structuring of the inquiry might well be done by a selected panel of system psychologists. Each respondent would be required to submit samples and extracts of several pages of his own task analysis in the form of working documentation. The utility of the entire inquiry would depend on the candor of the respondents and the extent to which each was able to link the task information he sought to some personnel hypothesis(es) or decision(s). This would be the test of relevance.

A Data-Oriented Empirical Approach

Drs. E. A. Fleishman and R. W. Stephenson (American Institutes for Research) have suggested a means of combining creative insights and hypotheses based on data with empirical tests of these hypotheses (5). The hypotheses center around the kinds of human performance relevant to a given set of findings; thus, they are aimed at creating task taxonomies. A typical classification objective for a selection instrument would be "to post-dict the relative performance of individuals in one specified task based on their relative performance in another specified task." The relevant criterion measure would be: "Can the provisional approach to classification be used to predict factor loadings and validity coefficients?" These hypotheses would be developed by experts in practical application of tests--not necessarily those most familiar with statistical data manipulations.

A somewhat different picture is applicable to human factors engineering. Presumably a tentative taxonomic structure has been developed for operational tasks. See, for example, that proposed for task structure by R. B. Miller (1), (6). Tasks would be defined by essential transactions, variables, and conditions. A given laboratory finding in the research literature generally implies some principle whereby a performance can be improved or hampered--assuming a high level of operator motivation. An expert in human engineering and performance would attempt to generalize the relevance of the finding to one or more members of a task family or taxonomic category. For example, it has been found that slight to moderate amounts of visual noise assist in some kinds of detection. Assuming this is, indeed, a finding, two steps could be taken. Hypotheses that this would hold true in other vigilance and detection tasks could be tested. And, hypotheses generalizing this finding could be tested for applicability to "identification" tasks, as defined in the tentative taxonomy. If the finding held within a class of tasks but had no systematic effect on another class of tasks, one basis for differentiation between classes of task according to a design factor would be established. There are some severe methodological difficulties in applying this approach as a basis for differentiation and assimilation of hypothetical task entities. Some behavioral principles apply to all tasks (e.g., fatigue after prolonged activity, rapidity in delivering knowledge of performance effect) so that empirical tests of relevance of behavioral findings may be more significant for the classification of the research literature than for

development of a task taxonomy. It is not logically necessary or even likely that one classification will be paralleled by the other. Despite these difficulties, the approach is worth exploring.

An alternative post-diction approach proposed by Drs. Fleishman, Teichner, & Stephenson (5) is to use decisions as a criterion measure, and to post-dict the nature of decisions after the decisions have already been made. One might, for example, obtain information about a selected decision (e.g., to subgroup instruments) in such a way that the characteristics of the task could be classified according to whatever taxonomy is being evaluated at the time. One could then evaluate the taxonomy in terms of its ability to post-dict the outcome of the selected decision. The outcome of the decision might be represented by a matrix of pluses (incremental benefit), minuses (decremental effect), and zeroes (no noticeable effect). If the taxonomy cannot post-dict the decision outcomes in ways that make sense to the experts, the odds are that the taxonomy is missing some key classifiers of significance to the eventual users of the taxonomy.

The following are major decisions that might be post-dicted in training.

1. What breakout of the total job can be made so that "part-task" (or "task") training can be effected apart from the rest of the job context, but permitting transfer of learning to the total job situation? The training efficiency (and cost saving) from so-called part-task training can be substantial. (There seems to be no proper expression to denote this idea; the expression "part-task training" is awkward and partially misleading.)

2. How to sample from the input variables that make up the operational universe of task stimuli and situations in order to "program" training content most effectively and efficiently. In this context, effectiveness has a transfer of training implication meaning a reliable applicability to the full universe of job situations. Efficiency is the rate at which a given level of training effectiveness is attained at a given level of cost per student.

Supporting, but secondary, factors consist of use of training devices and techniques (procedures) of one kind or another. Degree of fidelity of simulated displays, display programs, controls and display-control relationships has a traditional significance somewhat reduced in the age of plastics and programmable computers, but still meaningful in terms of taxpayer dollars.

Empirical studies have supported the hypothesis that where there is symbolic mediation of procedural tasks (such as in troubleshooting and other forms of deliberate decision making), the cognitive elements can be learned even though there are large differences between the learning stimulus and the operational stimulus. This difference, as an example,

suggests a legitimate distinction (at least for training purposes) between all the members of task families that operationally are performed with little or no cognitive mediation versus those that are primarily cognitive and symbolic. In other words, this would be a taxonomic distinction.

Experimental psychologists also have the problem of properly generalizing their findings not only with respect to behavior theory, but also with regard to the range of tasks (in the laboratory or in real life) to which the findings should apply. Traditionally, experimenters have been chary of explicit generalization to kinds of tasks. The particular inventive-empirical approach proposed by Drs. Fleishman, Teichner, and Stephenson (5) would utilize the literature to evaluate hypotheses about such generalizations. Simply stated, post-dictions would be made that two or more laboratory studies in the literature would have similar outcomes with respect to an experimental variable, such as a stressor, on performance. In this way, the professional literature could accelerate the rate at which a research taxonomy would develop. Provisional taxonomies could be continuously refined and extended until, perhaps, they became coextensive with behavior theory.

PRACTICAL EVALUATION OF TAXONOMIC DEVELOPMENT

The major thesis of this report is that a task taxonomy should be aimed at making or converting task descriptions that will assist in identifying and using psychological information (in one form or another) for making system design and personnel subsystem decisions. Task Taxonomy is therefore an information getting and decision making tool. As such, it must be evaluated as any tool is evaluated--by utilitarian criteria.

Information that leads to the choice of a given selection test or procedure is serving a design decision. This is also true of information that leads to the choice of a work-space configuration or to a training regimen. The application of a classification rubric to a collection of data adds information to those data. The process of relating a collection of statements or of data to a given decision--or class of decisions--adds information to those statements or data. In all these matters, the taxonomy serves as an information tool.

It should be emphasized that a taxonomy does not consist merely of a list of names. The substance of a taxonomy consists in the definitions accompanying the names--the instructions for proper use to some potential user. There is no intrinsic rule for the minimum amount of definitional context that should accompany the classificatory name and establish it as a principle of division and of extension. The definition may be as brief as a dictionary statement or as extended as a chapter in a book. Occam's razor does not apply to these definitions. Other things equal, of course, the more compact an instrument the better.

An adequate "evaluation" of a tool should result from sampling each of three interacting factors: the skill of the tool user, the properties of the tool itself, the kinds of subject matter or substance on which the tool is used. Operationally, a taxonomy is a procedure to assist in making decisions in classifying subsets of some universe of objects or events. Operationally, the classification decisions are valuable insofar as they promote some testable set of action decisions. In task taxonomy, I call these personnel subsystem design decisions.

An experimental evaluation, based on several kinds of pragmatic tests for tools, could be outlined as follows.

Prepare a course of instruction on a proposed task taxonomy for prospective system psychologists. (One might choose subsets of system psychologists--such as human factors specialists, training specialists, selection specialists--and then partition the course of instruction accordingly.) The instruction would aim at developing decision making skills with the taxonomic instrument <u>and</u> its supporting context (e.g., task descriptions and personnel subsystem decision structures).

Put the experimental students to work using the proposed taxonomy. Use the following criteria for measurement and comparison with a control group.

1. <u>How long does it take to learn, and how extensive are the prerequisites for learning to use the tool with some realistic criterion of utility?</u> This is a general measure of goodness of a tool. A comparison test might hold training time constant for experimental and control groups, and measure performance factors.

2. <u>Does use of the tool tend to rule out potentially valuable alternatives that might have been perceived without using the tool? On the other hand, does the tool's use open up alternatives and possibilities that otherwise would not have been considered?</u> The subjects would be required to specify hypotheses and evaluate them in one or more aspects of the design enterprise. Either the alternatives would be tested by implementation in practice (highly impractical) or experts would critique them.

3. <u>Does use of the tool (assuming user skill) tend to land the student in the right solution ballpark either in the empirical aspects of solving the problem--such as behavior predictions of a useful kind--or in finding relevant literature?</u> Relevant literature is that which contains data and/or design guidance which the student is able to identify by name or with the help of a psychological thesaurus. (Judgment of experts about relevance would be more practical than empirical tests.)

4. <u>Does use of the tool assist heuristically in homing towards improved solutions as one makes design interactions? The realist recognizes that design decisions are fundamentally iterative--no solution is</u>

quite right at the first moment of thinking about it. This implies that a good problem solving tool should aid both in convergent thinking and in divergent thinking. An operational test here would be the minimum number of empirical tests, amount of research facility and research costs required for a given goodness of operational results.

5. Does the problem solving tool enable programmatic assessment of progress towards the objectives? The goodness of the personnel subsystem design decisions made by the student would be evaluated by operational criteria. Evaluation would be a composite of measurements (or estimates) of the predictive goodness of the selection tests, the efficiency and effectiveness of training, the level of operational performance and reliability of the operator performing the missions, and so on. Since these factors interact, some difficulties in assigning appropriate weights of goodness would inevitably arise.

If a control group of students were taught an equivalent amount of time with any alternative set of concepts and procedures, and given similar problems to which similar criteria were attached, the outcomes in the form of profile scores from experimental and control subjects could be compared (hopefully taking into account interaction effects between individual differences and form of instruction).

In theory, this would be a measure of the goodness of a task taxonomy and the information structure and content it should support. In practice, making such an experimental comparison would be absurd.

There is another and ultimately more practical and pragmatic approach to evaluation. That is to count the number of individuals who, by some given date, use the tool in doing their work. Adoption (although partly a function of sales campaigns) is a function of the intrinsic worth of a product in concrete terms.

It is also possible to make localized and inconclusive tests of the predictive capabilities imparted by a system of classification to an "expert". Unfortunately for this approach, research findings may be as specific as the proper shape of the head of the indicator in a meter or as general as principles for any diagnostic search strategy. The probable result of this approach would be a taxonomy for human engineering equivalent to the total table of contents, plus index, of a human engineering handbook.

We seem to be left with pragmatic and essentially qualitative assessments of any proposed taxonomic tools, at least until some alternative taxonomies with specified use objectives can be compared experimentally with respect to these objectives (assuming the objectives can be quantified in comparable scales).

These comments are not intended to dead-end useful proposals for experimental evaluation of taxonomic tools for the personnel subsystem.

A tool can be highly useful without experimental proof of its value. The innovations in our culture introduced by the applications of the computer comprise a notable example. It may well be that, unlike the validation of discoveries in nature, inventions can be objectively evaluated only retrospectively by enumerating the things made possible by them.

A CRITIQUE OF SOME CURRENT LABORATORY APPROACHES

The development of a task taxonomy is a formidable quest. Assuming substantial facilities and attention are to be given to the enterprise, it would seem to be worth some patience in studying what can and cannot be done and reasons why.

A logical examination of some problems can show that at least some kinds of solutions are impossible, or so unreasonable in terms of underlying assumptions as to be virtually impossible of achievement or useful application. This conclusion may become apparent when the underlying assumptions are revealed, or when the methodological issues are exposed, or when the implications for applying some resulting product (such as a body of knowledge) are tested.

I have outlined below the major liabilities that I see in traditional laboratory research assumptions and procedures as they relate to development of a generalized task taxonomy for system design work.

Partitionable Entities

It may be useful conceptually to consider human performance as the product of a combination of functionally separable black boxes--like amplifiers, filters, generators--in the human organism, but they have dubious structural identification. A computer may achieve a given result by a large variety of different application programs that run the problem and control programs that operate the system. A switch stores information and a memory unit acts as a switch. The function exists in the program as much as in the wiring of the device; it exists as a succession of states as much as in the locus of its physical structure. The human structure itself seems to change with its patterns of experience--that is, learning "rewires the mechanisms".

The quest to abstract black box functions in the human seems bleak if not abortive. The exception may consist of the effector mechanisms--the subsystems which control and coordinate muscle behavior.

Nonsense Tasks

The bulk of the experimental literature centers around nonsense activities. Nonsense tasks are those in which the subject does not share

in the purpose for his activity and works in a stimulus-deprived situation. The exigencies of experimental control (and of consistency with other studies to be supported or refuted) require constraining the stimulus by the experimenter. The subject behaves, therefore, as a very specialized robot. These conditions are not generally representative of real-life human environments in which the operator acts as an intact specimen.

Abilities tests have, at least in degree, the same characteristic The subject is given a problem to solve with little of the context of real life situations. And scoring objectivity requires simple answers which must therefore exercise primarily "convergent" abilities. In real life, many alternative "answers" turn out to be equally good, and the answer may be a pattern of responses. The subject's capacity to develop strategies for a class of solutions (such as is the case if a task is repeated in mary missions and with variations in context) is given little or no opportunity to be manifest.

"Meaningful" tasks (according to any definition you choose) enable the subject to select and organize codes that give him mnemonic support in learning and in performance. Mnemonic structure may be the most significant aspect of learning real life tasks--and this opportunity is minimized in artifical stimulus-deprived situations. Mnemonic structure is the pattern of cognitive associations. In simplistic terms, when the operator is thinking of A in the context of doing C, he has a high probability of thinking of elements in an array, each element of which is the key to another array. We may call this a potential "train of associated ideas".

The designer of the traditional experiment is in a dilemma. If he cannot purify his independent and dependent variables he "won't know what he is measuring". If he extends the variable to a large number of situational contexts--including samples from real life tasks--his variance grows so large that the variable tends to disappear, and he has no conclusion to report, other than that the variable was swamped by "other factors". Large factorial studies are expensive, and even their mesh may be too loose for obtaining task structures applicable to <u>design</u> information.

Studies that seek to derive quantifications for information theory models may also have to tend to simplistic human activities. This is required by the need to measure the stimulus or input states and responses or output states in terms of discrete, countable information units. Doing so requires the experimenter's ability to encode and decode stimulus conditions and response conditions into terms enabling the assessment of "bandwidth" capabilities. Useful as these studies and the models they generate have been in some areas of human activity, their special requirements for quantifying data demand the equivalent of "nonsense tasks" or specialized kinds of tracking behavior. It is by no means clear that this is limited by the nature of the model, or by the aforementioned practical difficulties in dealing with complex patterns of stimulus and response.

Note, however, that "information" is not the equivalent of "meaning" in the usual sense of the word. Information theory deals with codes, and the relations between codes as signs, and their reference is not directly relevant to the theory. To the extent that human task learning and task performing has to do with the acquisition of meanings, or change in meaning, the information theory paradigms may seem unpromising if not sterile. (Please recall that various metholodologies are not being challenged here on their value to scientific knowledge, but on their probable utility for deriving a taxonomy.)

It is paradoxical that the requirements of scientific procedure in the laboratory tend to oppose those for developing a broad-based taxonomy of real world tasks. Research demands quantifications, control of variables, objective measurement, compatability with investigative materials used by colleagues. This forces abstractness of task and artifical simplicity in order that variables can be controlled both physically and statistically. It has been ironically observed that what can most readily be measured is likely to be of little utility in the non-laboratory world of complex events, interactions, and contingencies.

The artificiality of task situations in traditional research laboratories does not seem a fruitful base from which to develop a taxonomy. This is not to assert that, after a taxonomy has been developed, the results of many of these studies cannot serve useful purposes by being integrated and indexed according to appropriate task identities and class of design decision.

Inadequacy of Performance Data

Error data and error analysis can be the most fruitful kind of data from which to develop or modify behavioral principles. This has been true in academic as well as applied psychology. Attempts to interpret "failure mechanisms" have led to important discoveries in many fields.

Unfortunately, most empirical performance data, whether obtained from typists or from automobile drivers surviving accidents, is deficient in the identification of important circumstances--stimulus conditions and motive-incentive conditions. Kind and condition of error are usually inadequately characterized. Statistical summaries, however useful for actuarial purposes, have thrown away data about individual patterns of events that can be most trenchant for hypothesis formation.

Statistics about performance rates are generally too grossly clumped, and the distributions around means or medians are so large that they tend to be almost meaningless for predictive purposes in hypothetical conditions. This applies to rates ranging from typing productivity to programming productivity. That some typists can achieve occasional bursts of 18 or more keystrokes per second has indeed some value as an indication of

human limits. But conditions of selection, training, monitoring, input, environment, and procedure associated with hourly, daily and weekly throughput are of greater significance to personnel subsystem design enterprises. Performance data rarely are subsetted in ways that enable very useful analysis and generalization for predictive design purposes. The system designer is concerned with what will happen to performance if one or more of the parameters in the work configuration is <u>changed</u>. Without systematic data linked to individual cases, it is difficult or impossible to determine what factors performance is most sensitive to, assuming given levels in other factors.

Automatic sensors of human inputs and outputs fed into computer analysis hold promise for acquiring and processing much more information than was reasonable to do by eye and hand. But, the attachment of a sensor of a given kind imples a hypothesis by someone as to what is important to observe. Perhaps, ironically, the systematic development of such hypotheses may grow out of rather than produce a task taxonomic structure.

Publication of "No Difference Found" Data

It is as important for a consultant or applied behavior scientist to know in advance which factors make little or no difference, as which ones do. As every graduate student attempting an experimental dissertation knows to his anguish, it is difficult to frame experimental conditions that are more "significant" than individual variabilities. If the term "significant" conveys the criterion of practical differerce, in the applied field we find that motive-incentive conditions and procedure design generally blot out large ranges of difference in composites of other variables. But, researchers are motivated to avoid publishing "no difference found" studies as if they were failures; nor, to my knowledge, is a publication medium available for them.

The reporter of "no difference found" data may be less motivated to precision and completeness in describing the context in which the data were generated than is the reporter who is more likely to be subjected to the criticism of his colleagues. In principle, however, a Type II error creates just as serious a bias as a Type I error. This may be a region in which professional disciplines would have to develop and apply.

The Need for Change

It is this author's view that useful extrapolations cannot be made from meaningless to meaningful human tasks, that complex behavior in the real world is not composed of a mosaic of stimulus points linked to response points, and that the capability to respond to multiple streams of more or less concurrent series of signals is more than the sum of response to individual streams of signals from a given channel, source

or set of expectations. This view forms the background for the recommendations regarding laboratory approaches to studies about human tasks contained in the following section.

SOME POSITIVE RECOMMENDATIONS FOR A LABORATORY APPROACH

The problem solver uncovers and examines assumptions, and identifies objectives. He modifies both, as the study of assumptions reveals which objectives are realistic and which ones are not. He devises a strategy route for data collection that minimizes effort in reaching the objectives, or in reaching a decision that the objectives are unrealistic or not worth the trouble. The larger the research enterprise, the greater the importance of thorough examination of both assumptions and objectives.

A good strategy for reaching utilitarian objectives may not be equally good for reaching "scientific" objectives. A utilitarian objective is generally one that produces control of a phenomenon of established practical value. The efficient design of an effective personnel subsystem with low cost in time and resource is an example. Scientific objectives consist primarily of knowledge, and of control only as a by-product of knowledge. Knowledge of an entity that produces a disease is not equivalent to the control of the disease, although it may short-cut gaining such control. A good strategy for scientific objectives is one that maximizes the amount of knowledge acquired per unit of research effort, that is, per experiment conducted. A good theory is the most effective strategy for efficient collection of date about some domain of interest.

If one argues as I have that a task taxonomy makes sense only if it is conceived as a tool, then one's research strategy, if consistent, must be aimed at utilitarian objectives--tested by people using it for real purposes other than laboratory hypothesis-making and testing. Recommendations for a programmatic endeavor along these lines are set forth on the following pages.

Definition of Project Objectives

Project objectives serve as criteria for determining relative success of the product resulting from the effort. If taxonomy is a tool, the objectives should spell out in operational detail what decisions and operations it would support, and under what assumptions and limitations.

Here is an abbreviated example of a statement of objectives that might be used: A vocabulary of analytic-descriptive terms applicable to the observation of behavior samples, and a procedure for applying this vocabulary and its definitions such that a graduate student in applied psychology could learn and apply it to the universes of jobs A, B, C,...n

in a specified period of time. Criteria for "effective application" would be provided which should include the following operations: (a) use of the taxonomy as applied to a job situation to reference the literature and select "relevant" references, with explicit reasons for hypothesizing relevance; (b) use of the task/job description in making personnel subsystem design proposals and decisions of types to be explicitly identified; (c) prediction of outcomes of design alternatives based in part on the taxonomic descriptions, and specification of prediction criteria (which should be filled in according to reasonable aspirations); (d) specification of a strategy of efficient inquiry leading to generalizations in accordance with the structure and concepts in the taxonomy.

Project objectives need not be cast in steel. But, changes should continue to reference operational criteria of utility, either in the practical design situation or in the interests of more efficient scientific investigation.

Research Strategy

Progressive motion towards a defined goal through a large universe of alternative paths, possibilities and assumptions requires at minimum a loose strategy--in other words, some kind of explicit plan and a choice policy. Formulation of the plan depends, of course, on a goal definition. If the goal is changed, the change should be the outcome of rational choice among alternatives, rather than merely the abandonment of an inexpedient course of action that seems disappointing, or which becomes tiresome to the researcher.

A strategy consists of decision checkpoints at which alternatives may be considered. It also includes the relationships among parallel or complementary paths. Realistic planning should include one line of development/inquiry which, although less than an aspiration of the "ideal", has a high probability of utility. A tactical advantage in this development, in parallel, is that it provides a realistic base against which to evaluate the relative success and utility of the more ambitiously aimed work.

An example may clarify this point. Ideally, perhaps, a non-psychologist with a few hours of indoctrination would examine (by a method specified or unspecified) a verbal description of a job-task and environment (according to some specified format of description). Then by consulting a reference work or data bank with these rubrics, the analyst would make quantitative estimates of failure frequences--qualitative and quantitative--of some population of operators. This seems to be an unrealistic goal.

An objective far less ideal would be a product that consisted of a two-year training course for a graduate experimental psychologist. At

the end of this training he would be able to make an analysis, only some of which he could objectively verbalize, leading to (a) one or more man-machine system design hypotheses and (b) an experimental strategy that would enable, after a few hours of diagnostic testing, the determination of rough limits of quantitative and qualitative performance.

I doubt that the behavioral sciences have any examplers of this kind of strategy made explicit and communicable. There are real difficulties in implementation. Researchers are specialists, and they prefer to begin study of a behavioral problem from the corridors they know best--with phenomena and apparatus about which they feel comfortable. A complex objective, nevertheless, requires complex planning and a higher order of discipline than is needed for development/inquiry aimed at targets of opportunity.

Differentiating What Must Be Invented and What Must Be Discovered

The following comments apply equally to two contexts: flowcharting of the research/development plan or strategy; and, application of a research and development product. The term "invention" refers here to some act of judgment, expertise or creation. These acts may range from categories of task structure to decisions as to whether, for instance, photo interpretation falls into the category of "decoding" in the same sense that translating English into Russian is decoding, or written English text into typescript is decoding. Assume that the definition for decoding, however excellent and objective it may appear, does not explicitly include these examples.

The value systems of researchers strongly entrenched in the positivistic school lead them to emphasize whatever is empirically rooted in their work, leaving it to their critics to point out the semantic and other constructs in their structure of assumptions. The same applies to theory and to rationales, explicit or implicit, for the selection and naming of categories of data and the rejection of lines of inquiry.

It seems especially important to progress planned or completed in the exploratory phases of work on a problem--including that of <u>defining</u> the problem to solve--that differentiations be made carefully explicit between judgment and invention on the one hand and empirical findings on the other. If more than one approach is to be tried and compared at various stages of development, explicitness seems imperative.

Candor on this matter not only serves for better communication among participants (and competitors) in the enterprise. It should increase the systematic management of developmental work in the project by showing, among other things, what is relatively useless to try to verify empirically. This concept is applicable to research strategy.

A second differentiation between judgment/invention and empirical rigor derives from the use of the research product when completed. We

have stipulated that a given approach to taxonomic development is aimed at an end product. If the product is intended to have utility, procedures for the use of the product should be part of the product definition.

Procedures for using the product should specify the information necessary for carrying out each step in procedure. The basis for each necessary judgment should be stated. Observing behavior in vivo and describing it is a process of judgment in abstraction and in semantic operations. This is true, although perhaps to a lesser degree, when the observer must select from among a limited set of rubrics and verbal definitions. The judgment process becomes compounded when activities that are more or less concurrent must be identified, named and related (e.g., making a quantitative or qualitative prediction, applying a behavioral principle).

I recommend that a flowchart be prepared of the steps required in each of various kinds of research product application, and that the kind of human judgment required for each step be specified. A development advantage would derive from spotting such factors, which otherwise might appear to justify empirical studies of little value, no matter what outcome, because they contribute a negligible amount of variance to the total uncontrolled variance in the judgment process.

An approach to the development of a product should be preceded by a plan. If the plan is available, it should be possible to make the type of flowchart mentioned. An example of one such flowchart can be seen on the following page (see Figure 1).

Universe of Task Discourse

Programmatic enquiry and development should have some explicit subject matter boundaries, however crudely these may be expressed. Since the intent is to apply the task taxonomy to real life work and its environments, boundaries should be expressed in ways that can be referred at least roughly to examples of real jobs and work. A starting reference could be a dictionary of job titles or military job codes. From these, samples of task activities might be drawn almost at random, unless a more systematic procedure could be employed.

The gross definition of this task universe targeted for the prospective task taxonomy might very well be in layman's terms. The issue here is not what terminology would be employed, but the range of human operations in work contexts to which the taxonomy would apply.

If scope were limited to continuous tracking problems, for instance, a particular theoretical position, methodology, and laboratory setup could more readily be perceived as relevant to, but not necessarily inclusive of, the task universe. Were psychomotor tasks to define the target for examination, the factor analytic studies of Fleishman would [illegible] source of hypotheses. [illegible] essentially procedural tasks were

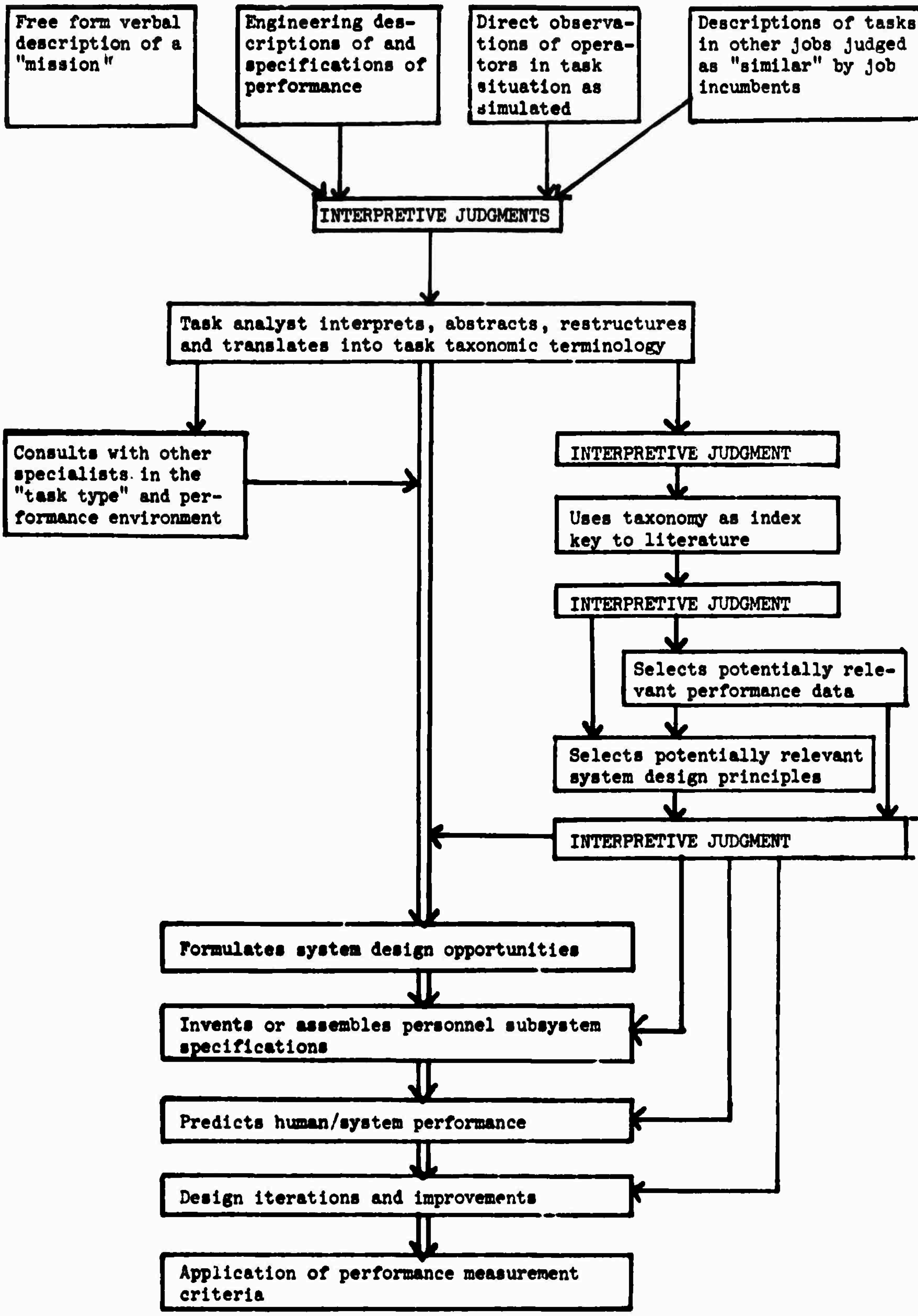

Figure 1. Role of interpretive judgment in the use of a task taxonomy and reference literature in system design

the subject of study and classification, then such concepts as Markov chains and matrices of conditional probabilities could appear to make sense. If mediating or cognitive activities were the targeted subject matter, Guilford's factor analytic studies could offer hypotheses. If the tasks were essentially defined in terms of interpersonal coordinations, personality inventories could be useful starting points for investigation.

In any event, if a subset of the total universe (even as now known) is to be tackled by a project, a rationale should be offered, together with proposals as to what should be done with the remainder of the real universe. As noted earlier, for pragmatic purposes it is quite conceivable that more than one set of taxonomic principles of division and clustering will be useful and even necessary. Description of the total universe of task discourse enables any subset of objectives to be perceived in perspective, and permits estimates of size of effort to be made.

Organizing the Findings and Implementing Design Recommendations

Review for a moment the critical steps involved in mapping out a research campaign, whether limited to one investigator or including vast numbers of them: definition of project objectives and explicit definition of boundaries of the universe of job-tasks relevant to project interest; formulation of a general research strategy which specifies priorities, policies, and criteria for further exploration or abandonment of a line of inquiry; creation of a flowchart distinguishing what must be "invented" from what must be discovered (by obtaining data confirming or denying hypotheses). A variation of the flowchart should diagram the elements in applying a result from any phase of work and kind of data collection endeavor to operational or design decisions. This flowchart should stipulate where and what kind of technical judgment is required in the application of a finding to a given kind of design decision or prediction.

The industrial community or the Department of Defense can readily provide sample problem situations for which task predictions or design recommendations need to be made. From these problems, those with characteristics that overlap some completed or partially completed area of examination (e.g., scanning and detection) may be selected as test cases. The test would be performed by individuals with characteristics specified by the project objectives as users applying the method of analysis, identification, and design recommendations. If several such individuals were used, a test of reliability among them would indicate the extent of objectivity of the procedures and terminology. As I have previously argued, this would not be equivalent to utility nor a substitute for it, only one desirable condition for it.

implementation of the design recommendations--even in the absence of trying out alternatives--would provide at least clinically useful support for the validity of the research effort. Difficulty encountered in converting the research development into an application would result in directing further work in the area.

This hand-in-glove arrangement between research leading to generalizable answers and the test of the utility of those answers in the real world seems the only assurance that the research effort will continue to have a practical payoff--practical in the very near future. Since the slow process of discovery of properties through the collection of data is supplemented by the relatively fast process of invention of tools, the entire operation can move forward more rapidly and on sounder grounds if invention and data collection proceed in tandem.

SPECIFIC SUGGESTIONS FOR LABORATORY STUDIES OF PERFORMANCE

There are three major questions generally asked about the operator's role in a system: Can he do the task at all? How well can he do it (i.e., with what reliability according to qualitative and quantitative criteria)? How much better or poorer can he do the job given a specific change in the environment, work-space design, or procedure? All three questions have to do with performance limits. (What the operator will do is, of course, a product of his motivation, skill, and work-task conditions.)

Data about the limits of performance in each of the task functions or categories according to the major transactional variables would provide a basis for answering these questions. Such data would also enable the task analyst to weight his examination of a real life job-task complex. A special focus of attention would be justified on a performance demand which seemed to approach some limit of a behavior capability. This and other directions for laboratory inquiry deserve enumeration.

Performance Limits on Task Function Variables

Laboratory investigation should be undertaken with tasks "meaningful" to the subjects. The same task variables should be embedded in at least three samples of task content: open-natural; maplike semirepresentational; symbolic. Degree of learning should extend to practice that is at least ten times the amount of practice taken by the subject to reach his plateau in performance on that task. The problem set should require some division of attention; the division of attention required should be a meaningful adjunct to the subject's primary task. (Examples include: sampling the status of a fuel indicator while piloting an aircraft near the end of its flight radius; attending to the paper supply indicator while using a copying machine; attending to the children

playing ball on a sidewalk flanking the road ahead while maneuvering in traffic).

Generalizing about the performance limit of a task function on the basis of data must be tentative if the prediction is to the task embedded in a complex of multiple-string activities.

Individual difference data should be kept during learning and during proficient performance of the task, for clinical examination and for correlational examination of within-task and between-task consistencies.

Some fuller explanation of the meaning of "meaningful tasks", activity groupings, and multiple-thread activity may be in order.

Meaningful tasks. Meaningful tasks are those in which the subject shares in purpose and criteria, has supportive information context, has initiative in developing strategies, encounters penalties for failure and satisfaction in perceived reward, has variable levels of aspiration, and has usually more than one effective goal route. There are usually criterion tradeoffs in meaningful tasks.

A meaningful task is also one in which the operations performed by the operator have a subjective mirroring in some form of imagery. The difference is exemplified by the series of terms L9F, 5VQ, TG3, in contrast to the series lake, swan, forest.

The versatility of the computer in displaying contexts of information to human subjects as well as in capturing and relating many aspects of their responses provides opportunity to control the richness both of task stimulus and of response measurement, thereby enabling study of meaningful tasks.

Activity groupings. Highly abstract or nonsense task materials that tend to prohibit any other than the most arbitrary groupings of activity likely to be a significant dimension of skill acquisition, should not be employed. It is evident, for example, that skilled typists translate words and phrases and "fields" from a source document into output patterns of movement; the transformation may include cognitive monitoring for "sense" in the text. They do not translate a series of alphanumeric characters into individual finger movements. Such activities introduce capabilities (and liabilities) that, from a predictive standpoint, differ from what would be learned from studying the transcribing of a series of random characters.

Task setups and task content which prohibit the formation of stimulus-response groupings, prevent development and observation of what may be one of the most significant aspects of real-life skills.

Multiple-thread task activity. Most if not all job-tasks require some division of attention. More than one thread of continuity needs to be sampled by the operator. Even a single apparent continuity may require division of attention between fresh input information to be processed and feedback from this output. Intellectual tasks also tend to have more than one "level" of continuity. Real life perceptual-motor activities require multiple strands of attention, at least on an intermittent basis. Skill in performing the job-task may often be demonstrated only in the act of balancing attention among a variety of ongoing activities sustained at the same time. Sustaining contrapuntal activity is certainly a test of short term memory competence which in turn demands a level of ability to handle the individual threads at a better level than "bare mastery".

Multiple-thread activity is involved in real life situations such as that posed to the automobile driver who must maintain a reference orientation to: his position in traffic with respect to other moving vehicles, the curb and other potential obstructions; potential moving obstructions such as a car quickly pulling out of a line of parked cars into his path; and, his location in a strange city of which he has only a maplike image; he must manage these continuities while searching for street signs and reading their content. Another example of multiple-string activity occurs when the driver, caught at an unexpectedly sharp curve while going faster than appropriate, must inhibit the powerful habit of applying brakes and instead deliberately maintain pressure on his accelerator. These phenomena have direct implications for training, cross-training, procedure design, human engineering design and, perhaps, for selection of operators.

A laboratory study on a phenomenon such as, "interpretation in a context of irrelevant stimuli" should include, as one control condition, the need to attend to a second string of ongoing activity. The information from the control may be more significant than that from the "pure" experimental condition; it may reveal the division-of-attention strategies adopted by the operator and his use of available initiatives.

Diagnostic Indicators that Limit Capability

Compared to that mentioned above, a less expensive and perhaps more valid approach for predictive purposes, which serves a supplemental purpose as well, would be to conduct interviews and observations aimed at determining factors that impose limits on work-task output. Real life tasks would be examined. The inquiry would be structured according to a set of task functions such as I have proposed. The method would consist of a combination of interview, observation of performing the task, and discussion in the course of walk-through of the job-task. The objective would be to find "indicators" that tend to limit capability in terms of human output. The output may be defined according to the criteria of rate, quality or reliability, or some combination. The

diagnostic indicator should be one which would enable physical simulation, selection and use of samples of human operators with identified characteristics to generate actual data on the simulator about that diagnostic indicator. These data should enable, within a reasonable range of error, predictive quantifications of performance in real work situations. Some examples of diagnostic indicators follow.

Electronic diagnosis of failure: "When I tried to figure out what next check to make, I often had trouble remembering what I'd already checked out as OK. Sometimes I'd have to make a series of tests and keep the test results in my head in order to know that the trouble wasn't in the sweep control". This is clearly a difficulty in short term memory, possibly complicated by absence of an appropriate diagrammatic representation and/or a diagnostic search strategy.

"My typing from the boss's manuscript is slowed because I have to stop now and then to correct the spelling of a word. I may even have to figure out what he was trying to say, and then change the wording." Here, straight coding transliteration rate collides with the need to identify and interpret so as to make changes in an input source. Multiple levels of input monitoring, decision-making and construction are involved in making corrections. Similarly, some proofreading editors report that trying to read for typographical errors and for sense at the same time makes for poor proofreading and is subjectively exhausting.

Dr. John Flanagan would recognize these examples as "critical incidents", and so they are. Collections of them, organized and classified by task function, task content, environment, and stage of learning (where applicable) would constitute a useful information source for predicting qualitative errors, as well as for potential skill delimiters. They could also be an educational gold mine for students of task analysis procedures who characteristically think of and observe only normative activities. And, they could be exercises for numerical solution by information theorists seeking to simplify analytic and predictive models of the function of this class of operations.

Qualitative Performance Errors

In this context, I shall define a performance error as unintentional, and as a failure to make a response which is in the operator's repertoire of capability (at least under some circumstances). Thus, the failure to detect or identify a sub-threshold cue is not regarded here as an "error", nor is an inability to make keystrokes at 20 cycles per second an "error".

Qualitative error information can be obtained in abundance by the critical incident method, by informal observation, and through more

systematic observation during experimental sessions under instrumental control. Because the "cause" of an error must be an inference, its nature can only suggest a hypothesis or imply a process or process variables.

For example, I believe that short term memory is a task function most susceptible to temporary deterioration under internal or external stressors (including fatigue), and second to this, scanning a field. This hypothesis would be tested most effectively by determining the kinds of errors made in complex task performances that depend in part on the normal integrity of these functions.

Assuming that the operator and observer share in the values associated with the goal criteria of a complex job-task, the kinds of error that occur may be indicative of the relative weight of a task function in the task complex. Thus, a large proportion of keystroke errors which are detected as such by the operator immediately upon making them (that is, before the next keystroke is made, or within two keystrokes) strongly suggests a motor interference, not a perceptual or mediating process error. The transposition of an entire word in running text suggests a failure in short term memory.

It should be recognized that what is a failure mechanism in one context may be an adaptive mechanism in another. Thus, object constancy has adaptive value in simplifying the information processing necessary to identify an object and holding its representation in mind; it is maladaptive when several objects are perceived as identical when they should be distinguished. What is "tunnel perception" in one situation is a concentration of regard in another. In the design of tasks for people or the selection of environments, knowledge of specific failure mechanisms associated with each task function in isolation and in concert with others enables greater sophistication and less trial and error in system design.

The knowledge and application of the knowledge of potential failure mechanisms associated with task function and task content, as well as stage of learning (since the mechanism may change with practice on the task), can be the most practical kind of information to provide for the design enterprise, and that bought at the lowest price.

Predicting the percentage <u>frequency</u> of given kinds of error is another matter. For an event which is relatively rare, extremely large samples must be used to obtain a stable frequency, and the prediction must be to collections of equally large number of cases. And of course, small constant influences may effect substantial changes in actuarial values. The aspiration to develop any data bank from which absolute frequencies of errors in new task contexts can be predicted seems to me to be at best an impractical enterprise.

"No Difference Found" Data

One objective for a taxonomy is a relatively small set of useful terms. Were all "no difference found" studies to be reported and suitably indexed, a major basis for grouping tasks and task contexts into sets and subsets might well be available. Unfortunately, the reports would have to identify the task characteristics in a way that would enable this clumping to be effected; and, this requirement imposes a need for judgment or an assumption of the availability of the very instrument to be created.

Insights into the behavior associated with common "tasks", in the layman's sense of the word, might be obtained from such data. For example, large numbers of studies on various type faces in typography have been generally negative on ease of reading the text printed in such type faces. In some cases it seemed incredible that a deformed typography could be read (by objective tests) as easily as the others. The point is, of course, that we do not read text letter by letter, but rather by word and phrase--by pattern and by contextual "meaning". Reading text is an interpretive process, or at least an identification of words, not of letters. Thus, the massive effects of redundancy overcome what might be local liabilities in a stimulus element. Within very broad ranges indeed, data on readability must depend on preferences rather than performance. This conclusion might be applied with peril to other code notations such as maps.

However, studies that show no differences among conditions permit the conclusion that the respective conditions (on the variable studied) can be treated as equivalent. Hence, a generalization can be made.

At this point it is necessary to distinguish a "no statistical difference found" from a "no practical difference found". I am referring to the "no practical difference" case, but accepting at least a loose statistical criterion of difference. A difference that may have no value in the world of practical affairs may be a highly important one for the world of science. A difference that might be useful for theory may be of no use whatever in the practical world, not because the phenomenon is not active in the real world as well as in the laboratory, but because its effects are swallowed up by other and more dominant variables or lost among interactions with them. I submit that data showing no difference (or only very little differences) among conditions could enable lumping together various sets of task rubrics, task contents, and task environments, depending on the identification of these factors in the experimental setups.

CONCLUDING STATEMENT

In this report I have advanced the view that a task taxonomy should be developed by invention rather than by scientific discovery. Taxonomies should be tested by practical utility as tools to be used by men, not by criteria of "truth" in the way a hypothesis about natural phenomena is tested experimentally. The taxonomy I envision is a human engineering creation that must be used with judgment, expertise, and uncertainty by the systems psychologist, whatever the context of his design decisions.

I have attempted to outline the kinds of work in the applied world of system design that a practical taxonomy could do. Nearly all of the kinds of problems and decisions described reflect some direct personal experience. I have also offered some proposals--those of others as well as my own--regarding practical steps for developing an applied taxonomy to be used as an information getting and decision-making tool, rather than as a rigorous "model" of human performance.

The major potential utility of a task description and methodology--or language--may be served to the extent that its use helps to structure and communicate or share problems in crisp terms of action and action alternatives, to anticipate trouble spots, and to record behavior in context. It may not be sufficient in itself for choosing from among decision alternatives in the quantitative sense.

I am quite critical of laboratory approaches to taxonomy development, primarily on the grounds that they are not adequately representative of the real world and do not lead to creation of useful tools. In spite of my reservations, I have tried to offer some constructive suggestions regarding important aspects of a laboratory approach to taxonomic development.

There are many behavioral scientists who would disagree with my comments. Among my colleagues there are dedicated researchers whose primary objective is to build psychological theory and who regard any practical fallout of their work with the indifference of the traditional scientist. For them, it appears that "prediction" does not mean specifying or forecasting human performance in the real world of multifold contaminant variables. Rather, it appears to mean a validated hypothesis strictly within the experimental laboratory milieu, usually as a follow-on to their own series of studies, subject matter, and controls. And the objective of a taxonomy seems equivalent to a structural extension of theory in which parsimony of terms is more important than decision-making utility in the world of work.

The "scientific-theoretical" orientation to taxonomic development i indeed quite different in objectives and evaluation criteria from that of the "user" orientation and empirical-inventive approaches proposed here. Regardless of which path proves correct in the long run, I contend that a user-oriented empirical-inventive approach is the best for our immediate needs. We cannot wait for the results of an approach oriented towards the discovery of some (as yet unknown) structural characteristics of performance. Tools are needed now to assist us in making system design decisions.

LITERATURE CITED

1. Miller, R. B. Development of a taxonomy of human performance: Design of a systems task vocabulary. Technical Report AIR-726/2035-3/71-TR11. Washington, D.C.: American Institutes for Research, March 1971. (U. S. Army Behavior and Systems Research Laboratory Research Study 71-4.)

2. Chapanis, A. The relevance of laboratory studies to practical situations. Ergonomics, 1967, 10(5), 557-577.

3. Reitman, W. What does it take to remember? In D. A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970. Chapter 14.

4. Bakan, D. On method. San Francisco: Jossey-Bass, 1967.

5. Fleishman, E. A., Teichner, W. H., and Stephenson, R. W. Development of a taxonomy of human performance: A review of second year's progress. Technical Progress Report 2. Washington, D. C.: American Institutes for Research, 1970.

6. Miller, R. B. Task description and analysis. In R. M. Gagné (Ed.), Psychological principles in system development. New York: Holt, Rhinehart & Winsonton, 1962. Chapter 6.

AMERICAN INSTITUTES FOR RESEARCH

Corporate Officers

John C. Flanagan, PhD
President

Brent Baxter, PhD
Executive Vice President—
Development

Edwin A. Fleishman, PhD
Senior Vice President

Paul A. Schwarz, PhD
Executive Vice President—
Operations

Board of Directors

John C. Flanagan, PhD, Chairman

Frederick B. Davis, EdD

Robert M. Gagné, PhD

Paul Horst, PhD

James M. Houston, LLB

S. Rains Wallace, PhD

Research Offices

ASIA/PACIFIC
Bangkok, Thailand
APO San Francisco 96346

KENSINGTON
10605 Concord Street
Kensington, Maryland 20795

PALO ALTO
P.O. Box 1113
Palo Alto, California 94302

PITTSBURGH
710 Chatham Center Office Building
Pittsburgh, Pennsylvania 15219

WASHINGTON
8555 Sixteenth Street
Silver Spring, Maryland 20910

Corporate Offices

PALO ALTO
P.O. Box 1113
Palo Alto, California 94302

PITTSBURGH
710 Chatham Center Office Building
Pittsburgh, Pennsylvania 15219

WASHINGTON
1775 K Street, N.W.
Washington, D.C. 20006